

STIC-Biotech/ChemLib

167494

119

From: Yu, Misook
Sent: Friday, September 30, 2005 4:58 PM
To: STIC-Biotech/ChemLib
Subject: 09/788,476

RECEIVED
OCT - 3 2005
STIC-BIOTECH DIVISION
(STIC)

Pls search SEQ ID NOs 1-3 including pending databases. Thank you,

Examiner Misook Yu, Ph.D.
571-272-0839 (Phone)
Art Unit 1642
REM-3A18 (Room)
REM-3C18 (Mail Box)

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 22:07:00 ; Search time 747.783 Seconds
(without alignments)
8254.755 Million cell updates/sec

Title: US-09-788-476A-1
Perfect score: 894
Sequence: 1 gggagtgagtgagggttaa.....taataaaaaaaatagaaaa 894

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/PCT_NEW PUB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09E_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq.*
- 23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
- 24: /cgn2_6/ptodata/1/pubpna/US11B_PUBCOMB.seq.*
- 25: /cgn2_6/ptodata/1/pubpna/US60_NEW PUB.seq.*
- 26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	100.0	894	9	US-09-788-476A-1
2	894	100.0	1154	9	US-09-788-476A-1
3	894	100.0	1154	14	US-10-091-483-90
4	873	97.7	873	9	US-09-788-476A-3
5	869.4	97.2	891	9	US-09-822-830A-389
6	833.8	93.3	1400	21	US-10-956-157-7827
7	833.8	93.3	1818	21	US-10-956-157-2592

8	733.4	82.0	3147	16	US-10-006-285-419	Sequence 419, App
9	690.8	77.3	2553	10	US-09-814-353-20673	Sequence 20673, A
c 10	595.2	66.6	2553	10	US-09-814-353-20673	Sequence 20673, A
11	498.4	55.7	1022	16	US-10-006-285-233	Sequence 233, App
12	460.8	51.5	488	17	US-10-242-535A-29658	Sequence 29658, A
13	460.8	51.5	488	18	US-10-085-783A-29658	Sequence 29658, A
14	458	51.2	470	17	US-10-242-535A-28856	Sequence 28856, A
15	458	51.2	470	18	US-10-085-783A-28856	Sequence 28856, A
16	431	48.2	470	17	US-10-242-535A-40257	Sequence 40257, A
17	431	48.2	470	18	US-10-085-783A-40257	Sequence 40257, A
18	389	43.5	426	17	US-10-242-535A-40571	Sequence 40571, A
19	389	43.5	426	18	US-10-085-783A-40571	Sequence 40571, A
20	379	42.4	379	17	US-10-242-535A-43248	Sequence 43248, A
21	379	42.4	379	18	US-10-085-783A-43248	Sequence 43248, A
22	336.2	37.6	463	10	US-09-918-995-15429	Sequence 15429, A
23	334	37.4	408	10	US-09-930-213-64	Sequence 64, Appl
c 24	281.4	31.5	5469	9	US-09-764-877-4000	Sequence 4000, Appl
25	281.4	31.5	5469	10	US-09-764-891-9371	Sequence 9371, Ap
c 26	281.4	31.5	5469	15	US-10-205-428-817	Sequence 817, App
c 27	281.4	31.5	5469	17	US-10-242-515-4000	Sequence 4000, Ap
c 28	281.4	31.5	9453	9	US-09-764-877-3999	Sequence 3999, Ap
c 29	281.4	31.5	9453	10	US-09-764-891-9370	Sequence 9370, Ap
c 30	281.4	31.5	9453	15	US-10-205-428-816	Sequence 816, App
c 31	281.4	31.5	9453	17	US-10-242-515-3999	Sequence 3999, Ap
c 32	246.2	27.5	300	16	US-10-006-285-196	Sequence 196, App
33	244	27.3	275	20	US-10-425-115-158622	Sequence 158622, A
34	211.8	23.7	255	10	US-09-930-213-531	Sequence 531, App
35	206	23.0	343	17	US-10-062-674-737	Sequence 737, App
36	174	19.5	257	17	US-10-062-674-1714	Sequence 1714, Ap
37	154.2	17.2	215	9	US-09-728-445-378	Sequence 378, App
38	154.2	17.2	215	22	US-10-964-549-378	Sequence 378, App
39	92.2	10.3	463	10	US-09-814-353-15759	Sequence 15759, A
c 40	55	6.2	399	22	US-10-450-763-26486	Sequence 26486, A
c 41	55	6.2	399	22	US-10-450-763-29210	Sequence 29210, A
c 42	54	6.0	5659	15	US-10-172-086-32	Sequence 32, Appl
c 43	54	6.0	5659	15	US-10-311-455-348	Sequence 348, App
c 44	54	6.0	5659	15	US-10-240-485-40	Sequence 40, Appl
c 45	54	6.0	5659	19	US-10-311-507-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-09-788-476A-1
; Sequence 1, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788,476A
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 894
; TYPE: DNA
; ORGANISM: human
US-09-788-476A-1

Query Match 100.0%; Score 894; DB 9; Length 894;
Best Local Similarity 100.0%; Pred. No. 1.8e-229;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GGGAGTGGAGTGGGGTAAACAGATCGGACCGAGCGGTGGAGCTCCATAAGCTAAAG	60
Db	1	GGGAGTGGAGTGGGGTAAACAGATCGGACCGAGCGGTGGAGCTCCATAAGCTAAAG	60
Qy	61	CTTCCCGAACTAAAGCAAGATGCTTCTGCTGTTGGAGACCAAGGGGATAAAGCAA	120
Db	61	CTTCCCGAACTAAAGCAAGATGCTTCTGCTGTTGGAGACCAAGGGGATAAAGCAA	120

```
Qy 121 GATCTTATCCAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAA 180
Db 121 GATCTTATCCAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAA 180
Qy 181 GAAGATGTACTGGAGATGAACAGAGAAAGAAACAAAGGCCATTGAGCTCCCTGTC 240
Db 181 GAAGATGTACTGGAGATGAACAGAGAAAGAAACAAAGGCCATTGAGCTCCCTGTC 240
Qy 241 AAAGAGGAAGAACCCCTGGAAGAACTGTTGATGTGCGACGACAGAGAAAGTGGTGAA 300
Db 241 AAAGAGGAAGAACCCCTGGAAGAACTGTTGATGTGCGACGACAGAGAAAGTGGTGAA 300
Qy 301 ATTACATCTGAATACCAACAGACTGAGAGATGAGAGAGGGCTGAACCAATCAATGTA 360
Db 301 ATTACATCTGNAATACCAACAGACTGAGAGATGAGAGAGGGCTGAACCAATCAATGTA 360
Qy 361 CCTGTGAGCTTGGAGATGAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTT 420
Db 361 CCTGTGAGCTTGGAGATGAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTT 420
Qy 421 CCAACAAAGGTCTGTCTATCTGATAACAAACCTTAACTTGGATAGCTGAAGGAA 480
Db 421 CCAACAAAGGTCTGTCTATCTGATAACAAACCTTAACTTGGATAGCTGAAGGAA 480
Qy 481 AGAGCTCAAAAGATTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
Db 481 AGAGCTCAAAAGATTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
Qy 541 AAATCTGAAAAAGAGAAAGGAGCGATTTGGGATTTGCAAGATTCAGCTGGAACCTGGAAC 600
Db 541 AAATCTGAAAAAGAGAAAGGAGCGATTTGGGATTTGCAAGATTCAGCTGGAACCTGGAAC 600
Qy 601 ACAGAGATACAGAGGCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 ACAGAGATACAGAGGCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy 661 AAAGTTCCTGATATCTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCT 720
Db 661 AAAGTTCCTGATATCTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCT 720
Qy 721 ATATATGCTTAAATGACAGTCATGTGCTACGTCCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 721 ATATATGCTTAAATGACAGTCATGTGCTACGTCCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Qy 781 CCCAGGTACATCATGAACTGCGGAGAGAGTTTGAATTTGACTTATGCTGTTTCAAGTTAAG 840
Db 781 CCCAGGTACATCATGAACTGCGGAGAGAGTTTGAATTTGACTTATGCTGTTTCAAGTTAAG 840
Qy 841 TTGTTGTTGTTTCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 894
Db 841 TTGTTGTTGTTTCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 894
```

```
RESULT 2
US-09-764-846-90
; Sequence 90, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ12
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1149)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-846-90
Query Match 100.0%; Score 894; DB 9; Length 1154;
Best Local Similarity 100.0%; Pred. No. 2.1e-229;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGAGTGGAGTGGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAGCTTAAAG 60
Db 177 GGAGTGGAGTGGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAGCTTAAAG 236
Qy 61 CTTGCCGAATTAAGCAAGAATGTCTTGTCTCGTGGTTTGGAGACCAGGGAATAAAGCAA 120
Db 237 CTTGCCGAATTAAGCAAGAATGTCTTGTCTCGTGGTTTGGAGACCAGGGAATAAAGCAA 296
Qy 121 GATCTTATCCAGACTCCAGGCATATCTTGAAGAAATGCTGTAAGAGGAGGCAATGAA 180
Db 297 GATCTTATCCAGACTCCAGGCATATCTTGAAGAAATGCTGTAAGAGGAGGCAATGAA 356
Qy 181 GAAGATGTACTGGAGATGAACAGAGGAAGAAACAAGCCCATTTGAGCTCCCTGTC 240
Db 357 GAAGATGTACTGGAGATGAACAGAGGAAGAAACAAGCCCATTTGAGCTCCCTGTC 416
Qy 241 AAAGAGAAAGAACCCCTCGAAACAACTGTTGATGTGCGACGAGAGAAAGTGGTGAA 300
Db 417 AAAGAGAAAGAACCCCTCGAAACAACTGTTGATGTGCGACGAGAGAAAGTGGTGAA 476
Qy 301 ATTACATCTGAATATACACAGACTGAGAGAAATGCGAGAGAGGCTGAAACGATTCATGTA 360
Db 477 ATTACATCTGAATATACACAGACTGAGAGAAATGCGAGAGAGGCTGAAACGATTCATGTA 536
Qy 361 CTTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGAGAGTGTGTTTGGGATTTCTTCAGTT 420
Db 537 CTTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGAGAGTGTGTTTGGGATTTCTTCAGTT 596
Qy 421 CCAACAAAAGGTCTGTCTATCTGATAACAAACCTATGTTTAACTTGGATTAAGCTGAAGGAA 480
Db 597 CCAACAAAAGGTCTGTCTATCTGATAACAAACCTATGTTTAACTTGGATTAAGCTGAAGGAA 656
Qy 481 AGAGCTCAAGATTTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
Db 657 AGAGCTCAAGATTTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 716
Qy 541 AAATCTGAAAAAGAGGAGAGGAGGATTTGGGATTTGTCACAAAGTTTCAAGTTTCAAGTCTG 600
Db 717 AAATCTGAAAAAGAGGAGAGGAGGATTTGGGATTTGTCACAAAGTTTCAAGTCTG 776
Qy 601 ACAGAGGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 777 ACAGAGGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 836
Qy 661 AAAGTTCCTGATATCTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCT 720
Db 837 AAAGTTCCTGATATCTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCT 896
Qy 721 ATATATGCTTAAATGACAGTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 897 ATATATGCTTAAATGACAGTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 956
Qy 781 CCCAGGTACATCCATGAACTGCGGAGAGAGTTTGAATTTGACTTATGCTGTTTCAAGTTAAG 840
Db 957 CCCAGGTACATCCATGAACTGCGGAGAGAGTTTGAATTTGACTTATGCTGTTTCAAGTTAAG 1016
Qy 841 TTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 894
Db 1017 TTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1070
```

```
RESULT 3
US-10-091-483-90
; Sequence 90, Application US/10091483
; Publication No. US20030049650A1
; GENERAL INFORMATION:
```



```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ12C1
; CURRENT APPLICATION NUMBER: US/10/091.483
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 348
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1149)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-091-483-90

```

Query Match	100.0%	Score	894	DB	14	Length	1154
Best Local Similarity	100.0%	Prod.	No.	2.1e-229			
Matches	894	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	GGGAGTGGAGTGAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAG	60				
Db	177	GGGAGTGGAGTGAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAG	236				
Qy	61	CTTCCCGAACTAAAGCAAGAAATGCTTGCTCGTGGTTTGGAGACAAGGGAAATAAAGCAA	120				
Db	237	CTTCCCGAACTAAAGCAAGAAATGCTTGCTCGTGGTTTGGAGACAAGGGAAATAAAGCAA	296				
Qy	121	GATCTTATCCACAGACTCCAGGCATATCTTCAAGAAACATGCTGAAGAGGAGGCCAAATGCAA	180				
Db	297	GATCTTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGGAGGCCAAATGCAA	356				
Qy	181	GAAGATGTACTGGGAGATGAAACAGAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTC	240				
Db	357	GAAGATGTACTGGGAGATGAAACAGAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTC	416				
Qy	241	AAAGAGGAAGAACCCCTTGAAAAAATCTTGATGTGGCAGCAGAGAGAAAAGTGGTGAAA	300				
Db	417	AAAGAGGAAGAACCCCTTGAAAAAATCTTGATGTGGCAGCAGAGAGAAAAGTGGTGAAA	476				
Qy	301	ATTACATCTGAAATACCAACACGACTCAGAGAATGCAGAGAGGGCTGAACAGATTCAATGTA	360				
Db	477	ATTACATCTGAAATACCAACACGACTCAGAGAATGCAGAGAGGGCTGAACAGATTCAATGTA	536				
Qy	361	CCTGTGAGCTTTGGAGAGTAAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTT	420				
Db	537	CCTGTGAGCTTTGGAGAGTAAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTT	596				
Qy	421	CCAAACAAAGTCTGTCTATCTGATPAAACAAACCTATGGTTTAACTTGGATTAAGCTGAGCAA	480				
Db	597	CCAAACAAAGTCTGTCTATCTGATPAAACAAACCTATGGTTTAACTTGGATTAAGCTGAGCAA	656				
Qy	481	AGAGCTCAAAGATTGGTTTGAATGCTCTCTTCAATCTCCAGAAAGCTCTGAAGATCATGAG	540				
Db	657	AGAGCTCAAAGATTGGTTTGAATGCTCTCTTCAATCTCCAGAAAGCTCTGAAGATCATGAG	716				
Qy	541	AAACTGAAAAAGAGGAAGGAGCGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGGAAACC	600				
Db	717	AAACTGAAAAAGAGGAAGGAGCGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGGAAACC	776				
Qy	601	ACAGAGGATACAGAGCCAAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCCCTGATGA	660				
Db	777	ACAGAGGATACAGAGCCAAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCCCTGATGA	836				
Qy	661	AAAGTTCTCTGATACATTTCTGTTCTCCAGTGTTTTCCATTTCTCTCTCTCTCTTGGTCAAC	720				
Db	837	AAAGTTCTCTGATACATTTCTGTTCTCCAGTGTTTTCCATTTCTCTCTCTCTTGGTCAAC	896				
Qy	721	ATATATGCTTAATGCACAGTCAATGTGCTACGTCTCTGCTCGCAATAGGGAGCATGTA	780				
Db	897	ATATATGCTTAATGCACAGTCAATGTGCTACGTCTCTGCTCGCAATAGGGAGCATGTA	956				

Qy	781	CCCCAGTACATCATGAAC	TGCGGCAGCAGT	TTGACTTATTCG	TGTTTCAGCTTTAAAGG	840
Db	957	CCCCAGGTACATCATGAAC	TGCGGCAGCAGT	TTGACTTATTCG	TGTTTCAGCTTTAAAGG	1016
Qy	841	TTGTTGTGTTTTTGGTTTT	TTCATTATGTTGCTT	TGTTAATAAAAAA	ATAAGAAA	894
Db	1017	TTGTTGTGTTTTTGGTTTT	TTCATTATGTTGCTT	TGTTAATAAAAAA	ATAAGAAA	1070

```

RESULT 4
US-09-788-476A-3
; Sequence 3, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788,476A
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PCR Product
US-09-788-476A-3

```

Query Match	97.7%	Score 873	DB 9	Length 873
Best Local Similarity	100.0%	Prod. No. 7.9e-224		
Matches 873	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	6	TGGAGTGGGGGTAA	CAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGC	65
DB	1	TGGAGTGGGGGTAA	CAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGC	60
QY	66	CGAACTAAAGCA	GAATGTCTTGCTGTGGTTTGGAGACCAAGGGATTAAGCAAGATCT	125
DB	61	CGAACTAAAGCA	GAATGTCTTGCTGTGGTTTGGAGACCAAGGGATTAAGCAAGATCT	120
QY	126	TATCCACAGACTCC	AGGCATATCTTGAAGACATGCTGAGAGGAGGCAATGAAGAAGA	185
DB	121	TATCCACAGACTCC	AGGCATATCTTGAAGACATGCTGAGAGGAGGCAATGAAGAAGA	180
QY	186	TGTACTGGGAGATG	AAACAGAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA	245
DB	181	TGTACTGGGAGATG	AAACAGAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA	240
QY	246	GGAGAAACCCCTG	AAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAANAATTAC	305
DB	241	GGAGAAACCCCTG	AAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAANAATTAC	300
QY	306	ATCTGAAATACCA	CGACTGAGAGAAATGTCAGAGAGGCGTGAAACGATTCATGTACCTGT	365
DB	301	ATCTGAAATACCA	CGACTGAGAGAAATGTCAGAGAGGCGTGAAACGATTCATGTACCTGT	360
QY	366	GAGCTTGGAGAGT	AAGAAAGCTGCTCGGCGACGTAGGTTTGGGATTTCTTCAGTTCCAAC	425
DB	361	GAGCTTGGAGAGT	AAGAAAGCTGCTCGGCGACGTAGGTTTGGGATTTCTTCAGTTCCAAC	420
QY	426	AAAAGGTCGTCA	TCTGNATACAAACCTATGGTTTAACTTTGATAGCTGAAGGAAGAGC	485
DB	421	AAAAGGTCGTCA	TCTGNATACAAACCTATGGTTTAACTTTGATAGCTGAAGGAAGAGC	480
QY	486	TCAAAGATTGGT	TTGATGTCTTCTCAATCTCCAGAAAGTCTGAAAGATGATGAGAACT	545
DB	481	TCAAAGATTGGT	TTGATGTCTTCTCAATCTCCAGAAAGTCTGAAAGATGATGAGAACT	540
QY	546	AAAAAGAGGAAG	GAGCGATTTGGGATTTGTCACAAGTTTACGTGGAACCTGGAACCAACAGA	605
DB	541	AAAAAGAGGAAG	GAGCGATTTGGGATTTGTCACAAGTTTACGTGGAACCTGGAACCAACAGA	600

Qy 178 GAAGAAGTGTACTGGGAGATGAACACAGAGGAAGAAACAAAGAGCCCATTTGAGCTCCCT 237
Db 673 GAAGAAGTGTACTGGGAGATGAACACAGAGGAAGAAACAAAGAGCCCATTTGAGCTCCCT 732
Qy 238 GTCAAAGAGGAAGAACCCCTTGAAGAACTGTTCATGTGGCAGCAGAGGAAGAAAGTGGTG 297
Db 733 GTCAAAGAGGAAGAACCCCTTGAAGAACTGTTCATGTGGCAGCAGAGGAAGAAAGTGGTG 792
Qy 298 AAAATTACATCTGAAATACCAACAGACTGAGAGAAATGCAAGAGGGCTGAAACGATTCAAT 357
Db 793 AAAATTACATCTGAAATACCAACAGACTGAGAGAAATGCAAGAGGGCTGAAACGATTCAAT 852
Qy 358 GTACTGTGAGCTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCA 417
Db 853 GTACTGTGAGCTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCA 912
Qy 418 GTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAG 477
Db 913 GTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAG 972
Qy 478 GAAAGAGCTCAAGATTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAAGATGAT 537
Db 973 GAAAGAGCTCAAGATTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAAGATGAT 1032
Qy 538 GAGAAATCTGAAAGAGAGAGGAGGATTTGGGATTTGTCAAGTTGAGCTGGAACTGGA 597
Db 1033 GAGAAATCTGAAAGAGAGAGGAGGATTTGGGATTTGTCAAGTTGAGCTGGAACTGGA 1092
Qy 598 ACCACAGAGGATACAGAGGCAAGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
Db 1093 ACCACAGAGGATACAGAGGCAAGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1152
Qy 658 TGAAGAGTTCCTGATACCTTTCTGTCTCTCCAGTGTTCCTTCTCTCTCTCTCTCTCTCTCT 717
Db 1153 TGAAGAGTTCCTGATACCTTTCTGTCTCTCCAGTGTTCCTTCTCTCTCTCTCTCTCTCTCT 1212
Qy 718 CACATATATGCTTAAATGCACAGTCATGTGCTTACGCTTCTGCTTCTGCTTCTGCTTCTGCT 777
Db 1213 CACATATATGCTTAAATGCACAGTCATGTGCTTACGCTTCTGCTTCTGCTTCTGCTTCTGCT 1272
Qy 778 GTACCCAGGTACATCCATGAACATGCGCAGCAGTGTGACTTATGCTTCTGCTTCTGCTTCTGCT 837
Db 1273 GTACCCAGGTACATCCATGAACATGCGCAGCAGTGTGACTTATGCTTCTGCTTCTGCTTCTGCT 1332
Qy 838 AGGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 894
Db 1333 AGGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1389

RESULT 7

US-10-956-157-2592
; Sequence 2592, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956.157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2592
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-2592

Query Match 93.3%; Score 833.8; DB 21; Length 1818;
Best Local Similarity 99.8%; Pred. No. 3.9e-213;
Matches 835; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 58 AAGCTTCCCGAAGCTAAAGCAAGAAATGTCTTCTCGTGGTTTGGAGACCAAGGGAATAAAG 117
Db 971 AAACTTGGCCGAAGCTAAAGCAAGAAATGTCTTCTCGTGGTTTGGAGACCAAGGGAATAAAG 1030
Qy 118 CAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAATGCTGAAAGAGGAGGCAAT 177
Db 1031 CAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAATGCTGAAAGAGGAGGCAAT 1090
Qy 178 GAAGAAGATGTACTGGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTTGAGCTCCCT 237
Db 1091 GAAGAAGATGTACTGGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTTGAGCTCCCT 1150
Qy 238 GTCAAAGAGAGAAACCCCTTGAAGAACTGTTCATGTGGCAGCAGAGAGAAAGTGGTG 297
Db 1151 GTCAAAGAGAGAAACCCCTTGAAGAACTGTTCATGTGGCAGCAGAGAGAAAGTGGTG 1210
Qy 298 AAAATTACATCTGAAATACCAACAGACTGAGAGAAATGCAAGAGGGCTGAAACGATTCAAT 357
Db 1211 AAAATTACATCTGAAATACCAACAGACTGAGAGAAATGCAAGAGGGCTGAAACGATTCAAT 1270
Qy 358 GTACTGTGAGCTTGGAGAGTAAGAAAGTGTCTGGGAGCTAGGTTTGGGATTTCTTCA 417
Db 1271 GTACTGTGAGCTTGGAGAGTAAGAAAGTGTCTGGGAGCTAGGTTTGGGATTTCTTCA 1330
Qy 418 GTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAG 477
Db 1331 GTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAG 1390
Qy 478 GAAAGAGCTCAAAGATTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAAGATGAT 537
Db 1391 GAAAGAGCTCAAAGATTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAAGATGAT 1450
Qy 538 GAGAAATCTGAAAG 597
Db 1451 GAGAAATCTGAAAG 1510
Qy 598 ACCACAGAGATACAGAGGCAAG 657
Db 1511 ACCACAGAGATACAGAGGCAAG 1570
Qy 658 TGAAGAGTTCCTGATACCTTTCTGTCTTCTCCAGTGTTCCTTCTCTCTCTCTCTCTCTCTCT 717
Db 1571 TGAAGAGTTCCTGATACCTTTCTGTCTTCTCCAGTGTTCCTTCTCTCTCTCTCTCTCTCTCT 1630
Qy 718 CACATATATGCTTAAATGCACAGTCATGTGCTTACGCTTCTGCTTCTGCTTCTGCTTCTGCT 777
Db 1631 CACATATATGCTTAAATGCACAGTCATGTGCTTACGCTTCTGCTTCTGCTTCTGCTTCTGCT 1690
Qy 778 GTACCCAGGTACATCCATGAACATGCGGAGCAGTGTGACTTATGCTGTTTTCAGCTTTTA 837
Db 1691 GTACCCAGGTACATCCATGAACATGCGGAGCAGTGTGACTTATGCTGTTTTCAGCTTTTA 1750
Qy 838 AGGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 894
Db 1751 AGGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1807

RESULT 8

US-10-006-285-419
; Sequence 419, Application US/10006285
; Publication No. US20030165854A1
; GENERAL INFORMATION:
; APPLICANT: Mary Jane Cunningham
; APPLICANT: Matthew R. Kaser
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS
; FILE REFERENCE: PA-0039 US
; CURRENT APPLICATION NUMBER: US/10/006.285
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PERL Program
; SEQ ID NO 419
; LENGTH: 3147

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165854A1 235169.27
US-10-006-285-419

Query Match      82.08; Score 733.4; DB 16; Length 3147;
Best Local Similarity 99.9%; Pred. No. 4.9e-106;
Matches 734; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 57 AAAGCTTCCGCAACTAAAGCAAGAAATCTTGTCTGCTGGTGTGGAGACCAAGGAATAAA 116
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 97 AAAGCTTCCGCAACTAAAGCAAGAAATCTTGTCTGCTGGTGTGGAGACCAAGGAATAAA 156
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 117 GCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGGCAAA 176
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 157 GCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGGCAAA 216
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 177 TGAGAGATGCTACTGGGAGATGAACACAGAGGAGAGAAACAAAGCCCATTCAGCTCCC 236
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 217 TGAAGAAGATGTACTGGAGATGAACACAGAGGAGAGAAACAAAGCCCATTCAGCTCCC 276
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 237 TGTCAAAGAGAGAAACCCCTGAAAAAATCTTGTATGTGCGACGACAGAGAAAGTGGT 296
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 277 TGTCAAAGAGAGAAACCCCTGAAAAAATCTTGTATGTGCGACGACAGAGAAAGTGGT 336
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 297 GAAATTTACATCTGAAATACACAGACTGAGAGAAATCGAAGAGGGCTGAACGATTCAA 356
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 337 GAAATTTACATCTGAAATACACAGACTGAGAGAAATCGAAGAGGGCTGAACGATTCAA 396
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 357 TGTACCTGTGAGCTTGAGAGTAGAAGCTGCTCGGGCAGCTAGGTTGGGATTTCTTC 416
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 397 TGTACCTGTGAGCTTGAGAGTAGAAGCTGCTCGGGCAGCTAGGTTGGGATTTCTTC 456
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 417 AGTTCCAAACAAAGCTGTGTCATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAA 476
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 457 AGTTCCAAACAAAGCTGTGTCATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAA 516
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 477 GGAAGAGCTCAAGATTTGGTTTGAATGTCTTTCAATCTCCAGAAAGTCTGAAGATGA 536
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 517 GGAAGAGCTCAAGATTTGGTTTGAATGTCTTTCAATCTCCAGAAAGTCTGAAGATGA 576
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 537 TGAGAAACTGAAAGAGGAGGAGCGATTTGGGATTTGTACAGTTTCAGCTGGAACCTGG 596
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 577 TGAGAAACTGAAAGAGGAGGAGCGATTTGGGATTTGTACAGTTTCAGCTGGAACCTGG 636
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 597 AACCACAGAGGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 656
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 637 AACCACAGAGGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 696
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 657 ATGAAAAGTTCCTGATATCTTCTCTTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTGG 716
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 697 ATGAAAAGTTCCTGATATCTTCTCTTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTGG 756
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 717 TCACATATATCCCTAAATGCAAGTCTATGTCCTACCTGCTGCTGCTGCTGCTGCTGCTGCT 776
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 757 TCACATATATCCCTAAATGCAAGTCTATGTCCTACCTGCTGCTGCTGCTGCTGCTGCTGCT 816
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 777 TGTACCCCGAGGTACA 791
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 817 TGTACCCCGAGGTACA 831
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-814-353-20673
; Sequence 20673, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
```

```
Qy 758 GCCTCGCAATGAGGAGCATGTACCCAGGTACA 791
Db 1375 GCCTCGCAATGAGGAGCATGTACCCAGGTCA 1408

RESULT 10
US-09-814-353-20673/c
; Sequence 20673, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20673
; LENGTH: 2553
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-20673

Query Match 66.6%; Score 595.2; DB 10; Length 2553;
Best Local Similarity 96.9%; Pred. No. 6.1e-149;
Matches 628; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

Qy 207 GGAAGAGAAACAAAGCCATTGAGTCCCTGTCAAGAGGAAGAACCCCTG-AAAAA 265
Db 695 GGAAGAGAAACAAAGCCATTGAGTCCCTGTCAAGAGGAAGAACCCCTGAAAAA 636

Qy 266 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATTTACATCTCAATACCAAGACTG 325
Db 635 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATTTACATCTCAATACCAAGACTG 576

Qy 326 AGAGAAATGCAAGAGGGCTCAACGATTCAATGTACCTGTGAGCTTGGAGAGTAAGAAAG 385
Db 575 AGAGAAATGCAAGAGGGCTCAACGATTCAATGTACCTGTGAGCTTGGAGAGTAAGAAAG 516

Qy 386 CTGTCGGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGCTGTCATCTGATA 445
Db 515 CTGTCGGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGCTGTCATCTGATA 456

Qy 446 ACAAACTTATGGTTAACTTGAATAGCTGAAGGAAGAGCTCAAGATTTGGTTGATG 505
Db 455 ACAAACTTATGGTTAACTTGAATAGCTGAAGGAAGAGCTCAAGATTTGGTTGATG 396

Qy 506 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAGAAAGAGGAGGAGCGAT 565
Db 395 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAGAAAGAGGAGGAGCGAT 336

Qy 566 TTGGATTTGTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTT 625
Db 335 TTGGATTTGTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTT 276

Qy 626 GGAAAGAGCAGAGCGCTTTGGGATTTGGGATTTGGGATTTGGGATTTGGGATTTGGGAT 685
Db 626 GGAAAGAGCAGAGCGCTTTGGGATTTGGGATTTGGGATTTGGGATTTGGGATTTGGGAT
```

```
Db 275 GGRAAAGAGCAGAGCGCTTTGGGATTTGGGATTTGGGATTTGGGATTTGGGATTTCT 216
Qy 686 CAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 745
Db 215 CAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 156
Qy 746 TGCCTACGTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACCTGGG 805
Db 155 TGCCTACGTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACCTGGG 96
Qy 806 CAGCAGTTTGACTTATTTGCTTTTACGCTTTAAAGTTTGTGTGTGTGTGTGTGT 853
Db 95 GCAGCAGTTGACTTA-TGCTGTTTTCAGCTTTAAAGTTGTGTGTGTGTGTGTGT 49

RESULT 11
US-10-006-285-233
; Sequence 233, Application US/10006285
; Publication No. US20030165854A1
; GENERAL INFORMATION:
; APPLICANT: Mary Jane Cunningham
; APPLICANT: Matthew R. Kaser
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS
; FILE REFERENCE: PA-0039 US
; CURRENT APPLICATION NUMBER: US/10/006,285
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PERL Program
; SEQ ID NO 233
; LENGTH: 1022
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165854A1 207354_Rn.3
; NAME/KEY: unsure
; LOCATION: 772-845
; OTHER INFORMATION: a, t, c, g, or other
US-10-006-285-233

Query Match 55.7%; Score 498.4; DB 16; Length 1022;
Best Local Similarity 87.3%; Pred. No. 3.9e-123;
Matches 580; Conservative 0; Mismatches 81; Indels 3; Gaps 3;

Qy 57 AAAGCTTTCGCAACTAAAGCAAGAATCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 116
Db 104 AAAGCTTTCGCAACTAAAGCAAGAATCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 163

Qy 117 GCAAGATCTTATCCACAGACT-CCAGGCATATCTTGAAGACATGCTGAAGAGAGGACAA 175
Db 164 ACAAGATCTTATCAATAGGCTACCCAGGCATATCTTGAAGACATGCTGAAGAGAGACAA 223

Qy 176 ATGAAGAAGATGTACTGGGAGATGAAACAGAGGAAGAAAGAAACAAAGCCATTCAGCTCC 235
Db 224 ATGAAGAAGATGTACTGGGAGATGAAACTGAGGAGAGAAACAAAGCCATTCAGCTCC 283

Qy 236 CTGTCAAAGAGGAGAAACCCCTCTGAAAGAACTGTGTGATGTGGCAGCAGAGAGAAAGTGG 295
Db 284 CTGTCAAAGAGGAGAAACCCCTCTGAAAGAACTGTGTGATGTGGCAGCAGAGAGAAAGTGG 343

Qy 296 TGAAGATTTACATCTGAAATACCAAGACTGAGAGATGCAAGAGAGGCTGAACGATTC 355
Db 344 TGAAGATTTACATCTGAAATACCAAGACTGAGAGATGCAAGAGAGGCTGAACGATTC 403

Qy 356 ATGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTG-GGCAGCTAGGTTTGGGATTTCT 414
Db 404 ATGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTGCGGCGAGGAGTTTGGGATTTCT 463

Qy 415 TCAGTTCCAAACAAAGGCTGCTGTCATCTGATAACAAACCTATGTTTAACTTGGATAAGCTG 474
Db 464 TCAGTTCCAAACAAAGGCTGCTGTCATCTGATAACCAAGGCAATGGTTTAACTTGGATAAGCTG 523

Qy 475 AAGGAAAGAGCTCAAGAGATTTGGTGTGAATGTCTCTTCAATCTCCAGAAAGCTCTGAAGAT 534
```

[illegible]

```

RESULT 12
US-10-242-535A-29658
; Sequence 29658, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Lew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIORITY APPLICATION NUMBER: US 10/085,783
; PRIORITY FILING DATE: 2002-02-28
; PRIORITY APPLICATION NUMBER: US 60/305,340
; PRIORITY FILING DATE: 2001-07-13
; PRIORITY APPLICATION NUMBER: US 60/275,017
; PRIORITY FILING DATE: 2001-03-12
; PRIORITY APPLICATION NUMBER: US 60/271,955
; PRIORITY FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 29658
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (37)..(37)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-29658

```

Db	241	CAAAAGAGAGAAACCCCTCGAATAAACTGTTGATGTGCACGACGAGAGAAAGTCGTGAA	300
Qy	300	AAATTACATCTCAAAATACACAGACTGAGAGAATGCAGAAAGAGGGCTGAAACGATTCAATGT	359
Db	301	AATACATCTCAAAATACCAAGACTGAGAGAATGCAGAAAGAGGGCTGAACGATTCAATGT	360
Qy	360	ACCTGTGAGCTTGGAGAGTAAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGT	419
Db	361	ACCTGTGAGCTTGGAGAGTAAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGT	420
Qy	420	TCCAAACAAAAGGCTGTGCATCTGATAACAAACCTATGGTTAACTTTGGATAAGCTGAAGGA	479
Db	421	TCCAAACAAAAGGCTGTGCATCTGATAACAAACCTATGGTTAACTTTGGATAAGCTGAAGGA	480
Qy	480	AAGAGCTC	487
Db	481	AAGAGCTC	488
RESULT 13			
US-10-085-783A-29658			
; Sequence 29658, Application US/10085783A			
; Publication No. US20040037841A1			
; GENERAL INFORMATION:			
; APPLICANT: ChondroGene Inc.			
; APPLICANT: Liew, C.C.			
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis			
; FILE REFERENCE: 4231/2002			
; CURRENT APPLICATION NUMBER: US/10/085.783A			
; CURRENT FILING DATE: 2002-02-28			
; PRIOR APPLICATION NUMBER: US 60/305,340			
; PRIOR FILING DATE: 2001-07-13			
; PRIOR APPLICATION NUMBER: US 60/275,017			
; PRIOR FILING DATE: 2001-03-12			
; PRIOR APPLICATION NUMBER: US 60/271,955			
; PRIOR FILING DATE: 2001-02-28			
; NUMBER OF SEQ ID NOS: 58994			
; SOFTWARE: PatentIn version 3.2			
; SEQ ID NO 29658			
; LENGTH: 488			
; TYPE: DNA			
; ORGANISM: Human			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (37)..(37)			
; OTHER INFORMATION: n is a, c, g, or t			
US-10-085-783A-29658			

```
Db 301 AACTACATCTGAATATACCACAGACTGAGAGAAATGCAGAAGGGCTGAACGATTCATGT 360
Qy 360 ACCTGTGAGCTTGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGT 419
Db 361 ACCTGTGAGCTTGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGT 420
Qy 420 TCCAAACAAAGGCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGTAAGCTGAAGGA 479
Db 421 TCCAAACAAAGGCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGTAAGCTGAAGGA 480
Qy 480 AAGAGCTC 487
Db 481 AAGAGCTC 488

RESULT 14
US-10-242-535A-28856
; Sequence 28856, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28856
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (23)..(23)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-28856

Query Match 51.2%; Score 458; DB 17; Length 470;
Best Local Similarity 99.8%; Pred. No. 1.9e-112;
Matches 469; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 3 GAGTGGAGTGAGGGGTAAACAAG-ATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGC 61
Db 1 GAGTGGAGTGAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGC 60
Qy 62 TTGCCGAACCTAAAGCAAGAATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAG 121
Db 61 TTGCCGAACCTAAAGCAAGAATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAG 120
Qy 122 ATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGAGCAATGAAG 181
Db 121 ATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGAGCAATGAAG 180
Qy 182 AAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCA 241
Db 181 AAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCA 240
Qy 242 AAGAGGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAAAGTGGTGAAGA 301
Db 241 AAGAGGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAAAGTGGTGAAGA 300
Qy 302 TTACATCTGAATATACCACAGACTGAGAGAAATGCAGAAGGGCTGAACGATTCATGTAC 361
Db 301 TTACATCTGAATATACCACAGACTGAGAGAAATGCAGAAGGGCTGAACGATTCATGTAC 360
Qy 362 CTGTGAGCTTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTC 421
Db 361 CTGTGAGCTTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTC 420
Qy 422 CAACAAAAGGCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAG 471
```

```
Db 301 TTACATCTGAATATACCACAGACTGAGAGAAATGCAGAAGGGCTGAACGATTCATGTAC 360
Qy 362 CTGTGAGCTTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTC 421
Db 361 CTGTGAGCTTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTC 420
Qy 422 CAACAAAAGGCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAG 471
Db 421 CAACAAAAGGCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAG 470

RESULT 15
US-10-085-783A-28856
; Sequence 28856, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28856
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (23)..(23)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-28856

Query Match 51.2%; Score 458; DB 18; Length 470;
Best Local Similarity 99.8%; Pred. No. 1.9e-112;
Matches 469; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 3 GAGTGGAGTGAGGGGTAAACAAG-ATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGC 61
Db 1 GAGTGGAGTGAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGC 60
Qy 62 TTGCCGAACCTAAAGCAAGAATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAG 121
Db 61 TTGCCGAACCTAAAGCAAGAATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAG 120
Qy 122 ATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGAGCAATGAAG 181
Db 121 ATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGAGCAATGAAG 180
Qy 182 AAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCA 241
Db 181 AAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCA 240
Qy 242 AAGAGGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAAAGTGGTGAAGA 301
Db 241 AAGAGGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAAAGTGGTGAAGA 300
Qy 302 TTACATCTGAATATACCACAGACTGAGAGAAATGCAGAAGGGCTGAACGATTCATGTAC 361
Db 301 TTACATCTGAATATACCACAGACTGAGAGAAATGCAGAAGGGCTGAACGATTCATGTAC 360
Qy 362 CTGTGAGCTTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTC 421
Db 361 CTGTGAGCTTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTC 420
Qy 422 CAACAAAAGGCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAG 471
```

Db 421 CAAACAAAGGCTGTCATCTGATACAAACCTATGGTTAACTTGGATAAG 470

Search completed: October 5, 2005, 05:55:38
Job time : 757.783 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:18:12 ; Search time 573.739 Seconds
(without alignment)
9224.147 Million cell updates/sec

Title: US-09-788-476A-1
Perfect score: 894
Sequence: 1 999agtgagtgagggttaa.....taataaaaaaaatagaaaa 894

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002as.*
7: geneseqn2002bs.*
8: geneseqn2003as.*
9: geneseqn2003bs.*
10: geneseqn2003cs.*
11: geneseqn2003ds.*
12: geneseqn2004as.*
13: geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	100.0	894	ABX13935	ABX13935 cDNA enco
2	894	100.0	1154	AAS29109	AAS29109 cDNA enco
3	894	100.0	1154	ABX68249	ABX68249 cDNA enco
4	894	100.0	1154	ADC25243	ADC25243 Human CDN
5	894	100.0	1520	ABN59623	ABN59623 Novel hum
6	891	99.7	1071	AAE88100	AAE88100 Human FLE
7	873	97.7	873	ABX13936	ABX13936 DNA enco
8	869.4	97.2	891	AAS62602	AAS62602 cDNA sequ
9	733.4	82.0	3147	ADL22609	ADL22609 Human liv
10	690.8	77.3	2553	ADL62461	ADL62461 Human ova
11	619	69.2	620	ACN39490	ACN39490 Tumour-as
12	595.2	66.6	2553	ADL62461	ADL62461 Human ova
13	498.4	55.7	1022	ADI22423	ADI22423 Rat liver
14	467.6	52.3	471	AAC00738	AAC00738 Human sec
15	451.6	50.5	558	ADP28791	ADP28791 Human adu
16	336.2	37.6	463	ACH28217	ACH28217 Human dif
17	334	37.4	408	AAH81555	AAH81555 Human imm
18	281.4	31.5	5469	AAK84119	AAK84119 Human met
19	281.4	31.5	5469	AAK37635	AAK37635 Human rep
20	281.4	31.5	5469	AAK06683	AAK06683 Human rep

C	21	281.4	31.5	5469	4	ABA08022	ABA08022 Human ova
C	22	281.4	31.5	5469	8	ABX60623	ABX60623 cDNA enco
C	23	281.4	31.5	5469	12	ADJ31373	ADJ31373 Human mus
C	24	281.4	31.5	9453	4	AAK84118	AAK84118 Human imm
C	25	281.4	31.5	9453	4	AAK37634	AAK37634 Human mus
C	26	281.4	31.5	9453	4	AAK06682	AAK06682 Human rep
C	27	281.4	31.5	9453	4	ABA08021	ABA08021 Human ova
C	28	281.4	31.5	9453	8	ABX60622	ABX60622 cDNA enco
C	29	281.4	31.5	9453	12	ADJ31372	ADJ31372 Human mus
C	30	281.4	31.5	9453	12	AAV89379	AAV89379 EST clone
C	31	246.2	27.5	300	10	ADI22386	ADI22386 Rat liver
C	32	231.2	25.9	591	2	AAH85623	AAH85623 Novel CDN
C	33	211.8	23.7	255	5	AAH82022	AAH82022 Rat diffe
C	34	196.8	22.0	232	8	AAH49101	AAH49101 Mouse DST
C	35	184.8	18.4	399	10	ADBS0756	ADBS0756 Primary r
C	36	154.2	17.2	215	6	ABX69315	ABX69315 Novel mur
C	37	92.2	10.3	463	5	ADL41869	ADL41869 Human ova
C	38	56	6.3	208765	12	ADQ97430	ADQ97430 Mouse can
C	39	55.4	6.2	1686	2	AAQ87587	AAQ87587 DNA enco
C	40	55	6.2	399	5	AAH90682	AAH90682 DNA enco
C	41	55	6.2	399	5	AAH93406	AAH93406 DNA enco
C	42	54.4	6.1	1997	13	ADR07778	ADR07778 Full leng
C	43	54	6.0	5659	6	ABL32375	ABL32375 Human imm
C	44	54	6.0	5659	6	ABL34487	ABL34487 Human met
C	45	54	6.0	5659	7	ADS99748	ADS99748 Complemen

ALIGNMENTS

RESULT 1
ABX13935
ID ABX13935 standard; cDNA; 894 BP.
XX
AC ABX13935;
DT 07-MAR-2003 (first entry)
XX
DE cDNA encoding novel human protein HCC-1.
XX
KW Human; hcc-1; human hepatocellular carcinoma tissue; HCC-M cell line;
KW cancer; hepatocellular carcinoma; antisense gene therapy; gene;
KW chromosome 7q22.1; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 25..657
FT /*tag= a
FT /product= "HCC-1"
XX
XX US2002107190-A1.
XX
XX 08-AUG-2002.
XX
XX 21-FEB-2001; 2001US-00788476.
XX
XX 25-FEB-2000; 2000US-0185116P.
XX
XX (CHUN/) CHUNG C M.
XX (CHAN/) CHAN L.
XX (OUKK/) OU K.
XX (ONGS/) ONG S.
XX (SEOW/) SEOW T K.
XX (LIAN/) LIANG C R.
XX (CHOO/) CHOONG M L.
XX (TANL/) TAN L K.
XX
XX Chung CM, Chan L, Ou K, Ong S, Seow TK, Liang CR, Choong ML;
XX Tan LK;
XX WPI; 2002-697878/75.
XX
XX P-PSDB; ABG72881.
XX
XX

CC activity). The nucleic acids and proteins display the following
CC activities Cytostatic, antibacterial, Virucide, Neuroprotective,

Query Match 100.0%; Score 894; DB 10; Length 1154;
Best Local Similarity 100.0%; Pred. No. 2.2e-230;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAGTGCAGTGGGGTAAACAAGATGCGGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 60
DB 177 GGGAGTGGAGTGAGGGGTAAACAAGATGCGGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 236
QY 61 CTTGCCGAACATAAGCAAGATGCTTCTGCTCGTGGTGGAGACCAAGGGAATAAAGCAA 120
DB 237 CTTGCCGAACATAAGCAAGATGCTTCTGCTCGTGGTGGAGACCAAGGGAATAAAGCAA 296
QY 121 GATCTTATCCACAGACTCAGGCAATCTTGAAGAACATGCTGAAGAGAGGCAAAATGAA 180
DB 297 GATCTTATCCACAGACTCAGGCAATCTTGAAGAACATGCTGAAGAGAGGCAAAATGAA 356
QY 181 GAAGATGTACTGGGAGTGAACAGAGGAAGAGAACAAAGCCCATTTGAGCTCCCTGTC 240
DB 357 GAAGATGTACTGGGAGTGAACAGAGGAAGAGAACAAAGCCCATTTGAGCTCCCTGTC 416
QY 241 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTAAA 300
DB 417 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTAAA 476
QY 301 ATTACATCTGAAATACCAAGACTGAGAGAAATGCAGAAAGGGCTGAAACGATTCAATGTA 360
DB 477 ATTACATCTGAAATACCAAGACTGAGAGAAATGCAGAAAGGGCTGAAACGATTCAATGTA 536
QY 361 CTTGAGCTGGAGAGTGAAGAGCTGCTGGGAGCTAGGTTGGGATTTCTTCAGTT 420
DB 537 CTTGAGCTGGAGAGTGAAGAGCTGCTGGGAGCTAGGTTGGGATTTCTTCAGTT 596
QY 421 CCAACAAAGTCTGTCTATCTGATAACAACTATGTTAACTTGGATTAAGCTGAAGGAA 480
DB 597 CCAACAAAGTCTGTCTATCTGATAACAACTATGTTAACTTGGATTAAGCTGAAGGAA 656
QY 481 AGAGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
DB 657 AGAGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 716
QY 541 AAATGAAAAAGAGAGGAGCGATTTGGGATTTGTCAAGTTCAAGTTCAAGCTGGAAC 600
DB 717 AAATGAAAAAGAGAGGAGCGATTTGGGATTTGTCAAGTTCAAGTTCAAGCTGGAAC 776
QY 601 ACAGAGGATACAGAGGCAAGAGGAAAGAGCGAGCGCTTTGGGATTTGCTGATGA 660
DB 777 ACAGAGGATACAGAGGCAAGAGGAAAGAGCGAGCGCTTTGGGATTTGCTGATGA 836
QY 661 AAAGTTCCTGATCTTTCTGTTCTCCAGTGTTCATTTCTCTCTTCTTCTTCTTCTGTCAC 720
DB 837 AAAGTTCCTGATCTTTCTGTTCTCCAGTGTTCATTTCTCTCTTCTTCTTCTTCTGTCAC 896
QY 721 ATATATGCTAAATGACAGTCAATGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 897 ATATATGCTAAATGACAGTCAATGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 956
QY 781 CCCAGGTACATCCATGAACTGCGGACAGCTTTGACTTATGCTTTGCTTTCAGCTTTAAGG 840
DB 957 CCCAGGTACATCCATGAACTGCGGACAGCTTTGACTTATGCTTTGCTTTCAGCTTTAAGG 1016
QY 841 TTGTTGCTTTTGTGTTTGTATGTTGCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 894
DB 1017 TTGTTGCTTTTGTGTTTGTATGTTGCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1070

RESULT 5
ABN59623
ID ABN59623 standard; cDNA; 1520 BP.
XX
AC ABN59623;

XX 28-JUN-2002 (first entry)
XX Novel human coding sequence SEQ ID NO: 34.
XX Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
XX antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
XX neuroprotective; antiparkinsonian; protein therapy; EST;
XX expressed sequence tag; gene; ss.
OS Homo sapiens.
XX WO200222660-A2.
XX 21-MAR-2002.
XX 10-SEP-2001; 2001WO-US026015.
XX 11-SEP-2000; 2000US-00659671.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX P-PSDB; ABB97210.
XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis.
XX Claim 1; SEQ ID NO 34; 509pp; English.
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
XX and cancer, to screen for drugs, to treat inflammatory conditions e.g.
XX rheumatoid arthritis, and to treat nervous system disorders e.g.
XX parkinson's disease. The present sequence is a coding sequence of the
XX invention
XX
XX Sequence 1520 BP; 442 A; 310 C; 423 G; 345 T; 0 U; 0 Other;
Query Match 100.0%; Score 894; DB 6; Length 1520;
Best Local Similarity 100.0%; Pred. No. 2.5e-230;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGAGTGCAGTGGGGTAAACAAGATGCGGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 60
DB 612 GGGAGTGCAGTGGGGTAAACAAGATGCGGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 671
QY 61 CTTGCCGAACATAAGCAAGATGCTTCTGCTCGTGGTGGAGACCAAGGGAATAAAGCAA 120
DB 672 CTTGCCGAACATAAGCAAGATGCTTCTGCTCGTGGTGGAGACCAAGGGAATAAAGCAA 731
QY 121 GATCTTATCCACAGACTCAGGCAATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAA 180
DB 732 GATCTTATCCACAGACTCAGGCAATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAA 791
QY 181 GAAGATGTACTGGGAGATGAAACAGAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTC 240
DB 792 GAAGATGTACTGGGAGATGAAACAGAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTC 851
QY 241 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTAAA 300
DB 852 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTAAA 911
QY 301 ATTACATCTGAAATACCAAGACTGAGAGAAATGCAGAAAGGGCTGAAACGATTCAATGTA 360

Db 912 ATTACATCTGAATACACACAGACTGAGAGAAATGAGAGAGGGCTGAACGATTTCAATGTA 971
Qy 361 CTTGTGAGCTTGGAGAGTAAAGAGCTGCTGGGCGAGCTAGGTTTGGGATTTCTTCAGTT 420
PI |||||
Db 972 CTTGTGAGCTTGGAGAGTAAAGAGCTGCTGGGCGAGCTAGGTTTGGGATTTCTTCAGTT 1031
XX |||||
Qy 421 CCACAAAAGGCTGCTCATCTGATACAAACCTATGTTAACTTGGATAGCTGAAGGAA 480
Db 1032 CCACAAAAGGCTGCTCATCTGATACAAACCTATGTTAACTTGGATAGCTGAAGGAA 1091
Qy 481 AGAGCTCAAAAGATTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
Db 1092 AGAGCTCAAAAGATTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 1151
Qy 541 AAACCTGAAAAGAGGAAAGGAGCGGATTTGGGATTTGTCAAAAGTTTCAGCTGGAATCGGAAC 600
Db 1152 AAACCTGAAAAGAGGAAAGGAGCGGATTTGGGATTTGTCAAAAGTTTCAGCTGGAATCGGAAC 1211
Qy 601 ACAGAGATACAGAGGCAAAAGAGAGGAAAGAGAGAGCGCTTTGGGATTTGCTGATGA 660
Db 1212 ACAGAGATACAGAGGCAAAAGAGAGGAAAGAGAGAGCGCTTTGGGATTTGCTGATGA 1271
Qy 661 AAAGTTCTCGATACCTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTTCTTCTTGGTCA 720
Db 1272 AAAGTTCTCGATACCTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTTCTTCTTGGTCA 1331
Qy 721 ATATATGCTTAAATGACAGCTATGTCCTACGTCCTGCTCGCAATGAGGAGCATGTA 780
Db 1332 ATATATGCTTAAATGACAGCTATGTCCTACGTCCTGCTCGCAATGAGGAGCATGTA 1391
Qy 781 CCCAGGTACATCCATCACTGCGGAGCGATTTGACTTATGCTGTTTTCAGCTTTTAAGG 840
Db 1392 CCCAGGTACATCCATCACTGCGGAGCGATTTGACTTATGCTGTTTTCAGCTTTTAAGG 1451
Qy 841 TTGTTGTGTTTCTTTTGTGATTGTTGCTTGTGTTAATAAAAAAATAGAAA 894
Db 1452 TTGTTGTGTTTCTTTTGTGATTGTTGCTTGTGTTAATAAAAAAATAGAAA 1505

RESULT 6
AAC88100
ID AAC88100 standard; cDNA; 1071 BP.
XX
AC AAC88100;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human FLEXHT-31 nucleotide sequence SEQ ID NO:86.
XX
KW Human; FLEXHT; full-length molecules expressed in human tissue;
KW diagnosis; gene expression; genetic linkage; genetic variability;
KW antianaemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;
KW cytosolic; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;
KW anti-gout; neuroprotective; antiarthritic; osteopathic;
KW antiproliferative; antirheumatic; antitumor; gene therapy; anaemia; gout;
KW epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;
KW cancer; immunological disorder; asthma; bronchitis; cirrhosis;
KW Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;
KW osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;
KW ulcerative colitis; ss.

XX Homo sapiens.
OS
XX WO200070047-A2.
XX
XX 23-NOV-2000.
XX
XX 12-MAY-2000; 2000WO-US013299.
XX
XX 14-MAY-1999; 99US-00311894.
XX 14-MAY-1999; 99US-00311937.
XX 14-MAY-1999; 99US-00311940.
XX

PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Tang YT, Lal P, Reddy R, Batra S, Baughn MR, Yang J;
PI Azimzai Y, Lu DAM, Au-Young J, Shih LL;
XX
DR WPI; 2001-016234/02.
DR P-PSDB; AAB36609.
XX
PT Human FLEXHT protein and DNA sequences, useful for treating immunological
PT disorders, developmental disorders, and cancers.
PS
PS Claim 5; Page 154; 168pp; English.
XX
XX AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules
CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The
CC present invention describes an isolated polypeptide (A) comprising an
CC amino acid sequence selected from one of 55 amino acid sequences 42-876
CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %
CC identical sequence, and a biologically active or immunogenic fragment of
CC the sequence. The FLEXHT proteins can have antianaemic, anticonvulsant,
CC antiarteriosclerotic, immunomodulatory, cytosolic, antiasthmatic,
CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antitumor, and
CC neuroprotective, antiarthritic, osteopathic, antiproliferative, antitumor and
CC antirheumatic activities, and can be used in gene therapy. The
CC polynucleotide sequences can be used to express the protein sequences.
CC Pharmaceutical compositions comprising FLEXHT can be used to treat
CC diseases or conditions associated with altered expression of functional
CC FLEXHT. The proteins and polynucleotides can be used to diagnose and
CC treat disorders including anaemia, epilepsy, arteriosclerosis,
CC atherosclerosis, developmental disorders, cancers, and immunological
CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,
CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,
CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and
CC ulcerative colitis
XX
SQ Sequence 1071 BP; 338 A; 190 C; 297 G; 246 T; 0 U; 0 Other;

Query Match 99.7%; Score 891; DB 5; Length 1071;
Best Local Similarity 100.0%; Pred. No. 1.4e-229;
Matches 891; Conservative 0; Mismatches 229; Indels 0; Gaps 0;
Qy 4 AGTGGAGTGGGGTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTT 63
Db 163 AGTGGAGTGGGGTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTT 222
Qy 64 GCCGAACCTAAAGCAAGAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123
Db 223 GCCGAACCTAAAGCAAGAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 282
Qy 124 CTTATCCACAGACTCCAGGCATATCTTGAAGACATGCTGAGAGGAGGCGCAATGAAGAA 183
Db 283 CTTATCCACAGACTCCAGGCATATCTTGAAGACATGCTGAGAGGAGGCGCAATGAAGAA 342
Qy 184 GATGTAAGTGGAGATGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 243
Db 343 GATGTAAGTGGAGATGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 402
Qy 244 GAGG 303
Db 403 GAGG 462
Qy 304 ACATCTGAATATCCACAGACTGAGAGATGTCAGAGAGGCGCTGAACGATTTCAATGTACCT 363
Db 463 ACATCTGAATATCCACAGACTGAGAGATGTCAGAGAGGCGCTGAACGATTTCAATGTACCT 522
Qy 364 GTGAGCTTGGAGAGTAAAGAGCTGCTCGGGAGCTAGGTTTGGGATTTCTTCAGTTTCCA 423
Db 523 GTGAGCTTGGAGAGTAAAGAGCTGCTCGGGAGCTAGGTTTGGGATTTCTTCAGTTTCCA 582
Qy 424 ACAAAGGCTGCTCATCTGATTAACAAACCTATGTTTAACTTGGATTAAGCTGAAGGAGAA 483
Db 583 ACAAAGGCTGCTCATCTGATTAACAAACCTATGTTTAACTTGGATTAAGCTGAAGGAGAA 642

Db	781	GGTACATCCATGAACCTGGCGACGAGTTGACATTATGCTGTTTTCAGCTTTAAGGTTGTT	840
Qy	846	GTGTTTTTGGTTTGGATTATGTTGCTTGTAAAT	878
Db	841	GTGTTTTTGGTTTGGATTATGTTGCTTGTAAAT	873
RESULT 8			
AD	AAS62602/c		
ID	AAS62602 standard; cDNA; 891 BP.		
AC	AAS62602;		
DT	14-FEB-2002 (first entry)		
XX	cDNA sequence #389 encoding novel human secreted protein.		
XX	Human secreted protein; hyperproliferative disorder; autoimmune disorder;		
KW	immune deficiency disorder; blood disorder; inflammatory disorder;		
KW	infectious disorder; gene therapy; antimicrobial; hepatotropic;		
KW	immunosuppressive; antirheumatic; ss.		
XX	Homo sapiens.		
OS	WO200177291-A2.		
PN	18-OCT-2001.		
XX	29-MAR-2001; 2001WO-US010485.		
XX	06-APR-2000; 2000US-0195604P.		
PR	(GEMY) GENETICS INST INC.		
PA	Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;		
PI	Gulukota K, Graham JR;		
XX	WPI; 2002-010900/01.		
DR	New polynucleotides encoding secreted proteins useful for treating e.g.		
PT	asthma, HIV and Crohn's disease.		
XX	Claim 1; Page 280; 391pp; English.		
XX	The present invention relates to the isolation of novel cDNA sequences		
CC	which encode human secreted proteins. The cDNA sequences have been		
CC	derived from a variety of human tissues. The invention also provides a		
CC	method for producing proteins from these polynucleotide sequences. The		
CC	proteins are useful for identifying compounds that modulate their		
CC	activity and production, and the cell is also useful for identifying		
CC	compounds that modulate expression of the polynucleotide sequences		
CC	encoding the secreted proteins. The sequences of the invention are useful		
CC	for treating diseases such as hyperproliferative disorders (e.g. cancer),		
CC	immune deficiency disorders (e.g. severe combined immunodeficiency		
CC	(SCID), autoimmune disorders (e.g. multiple sclerosis), blood disorders		
CC	(e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and		
CC	infectious disorders (e.g. hepatitis). The polynucleotide sequences of		
CC	the invention are also useful in gene therapy. AAS62214-AAS62838		
CC	represent the cDNA sequences of the invention that encode for novel human		
CC	secreted proteins		
XX	Sequence 891 BP; 221 A; 224 C; 157 G; 289 T; 0 U; 0 Other;		
SQ	Query Match 97.2%; Score 869.4; DB 6; Length 891;		
	Best Local Similarity 99.8%; Pred. No. 8.5e-224;		
	Matches 881; Conservative 0; Mismatches 1; Indels 1; Gaps 1;		
Qy	13	AGGGTTAAAGATGGCGACGCGAGACGGTGGAGCTCCATAGCTTAAGCTTGGCGAACTA	72
Db	886	AAGGGTTAAAGATGGCGACGCGAGACGGTGGAGCTCCATAGCTTAAGCTTGGCGAACTA	827
Qy	73	AAGCAAGAATGTCTTGCTGCTGGTTTGGAGACCAAGGGAATAAGCAAGATCTTATCCAC	132


```
Qy 336 GAAGAGGCTGAACGATTCAATGTACCTGTGAGCTTGAGAGTAAGAAAGCTGCTCGGGC 395
Dy 61 GAAGAGGCTGAACGATTCAATGTACCTGTGAGCTTGAGAGTAAGAAAGCTGCTCGGGC 120
Qy 396 AGCTAGGTTTCGGGATTTCTTCAGTTCCAAACAAAAGGCTGTCATCTGATACAAACCTAT 455
Dy 121 AGCTAGGTTTCGGGATTTCTTCAGTTCCAAACAAAAGGCTGTCATCTGATACAAACCTAT 180
Qy 456 GGTAACTTGATGAAGCTGAAGGAAAGAGCTCAAGATTTGGTTTGAATGCTCTTCAAT 515
Dy 181 GGTAACTTGATGAAGCTGAAGGAAAGAGCTCAAGATTTGGTTTGAATGCTCTTCAAT 240
Qy 516 CTCAGAAAGTCTGAAGATGATGAAACTGAAAGAGGAAAGAGCGATTTGGGATTTGT 575
Dy 241 CTCAGAAAGTCTGAAGATGATGAAACTGAAAGAGGAAAGAGCGATTTGGGATTTGT 300
Qy 576 CACAAGTTCAGCTGGAAGCTGAAACACACAGAGGATACAGAGGCAAGAGGAAAGAGC 635
Dy 301 CACAAGTTCAGCTGGAAGCTGAAACACACAGAGGATACAGAGGCAAGAGGAAAGAGC 360
Qy 636 AGAGCGCTTTGGGATTTGCTGATGAAAGTTCCCTGATACCTTTCTGTTCTCCAGTGTTC 695
Dy 361 AGAGCGCTTTGGGATTTGCTGATGAAAGTTCCCTGATACCTTTCTGTTCTCCAGTGTTC 420
Qy 696 CATTTCTCTCTTCTTTGCTGATATATATGCTTAATGACAGTCAATGCTACGTC 755
Dy 421 CATTTCTCTCTTCTTTGCTGATATATATGCTTAATGACAGTCAATGCTACGTC 480
Qy 756 CTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGACTGGCGAGCAGTTG 815
Dy 481 CTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGACTGGCGAGCAGTTG 540
Qy 816 ACTTATGCTGTTTTCAGCTTTAAGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 875
Dy 541 ACTTATGCTGTTTTCAGCTTTAAGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 600
Qy 876 AATAAAAAAATAGAAA 894
Dy 601 AATAAAAAAATAGAAA 619
```

RESULT 12

ADL62461/c

ID ADL62461 standard; DNA; 2553 BP.

XX

AC ADL62461;

XX

DT 20-MAY-2004 (first entry)

XX

XX Human ovarian cancer DNA marker #20673.

XX

KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.

XX

OS Homo sapiens.

XX

EN WO200170979-A2.

XX

PD 27-SEP-2001.

XX

PF 21-MAR-2001; 2001WO-0009126.

XX

PR 21-MAR-2000; 2000US-0191031P.

XX

PR 25-MAY-2000; 2000US-0207124P.

XX

PR 15-JUN-2000; 2000US-0211940P.

XX

PR 07-JUL-2000; 2000US-0216820P.

XX

PR 25-JUL-2000; 2000US-0220661P.

XX

PR 21-DEC-2000; 2000US-0257672P.

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Lee J, Lillie J;

XX

WPI; 2001-611502/70.

Novel isolated nucleic acid molecules (markers) overexpressed in ovarian cancer cells as compared to their normal non-cancerous ovarian cells are used to characterize stage, grade, histological type of ovarian cancer.

Disclosure; SEQ ID NO 20673; 106pp; English.

The invention relates to nucleic acid markers which are overexpressed in ovarian cancer cells as compared to their expression in normal (i.e. non-cancerous) ovarian cells. The invention also relates to polypeptides encoded by the markers, antibodies that selectively bind to the polypeptides, a method of inhibiting ovarian cancer in a patient at risk of developing ovarian cancer involving inhibiting expression of a gene corresponding to a marker of the invention and a method of treating a patient afflicted with ovarian cancer comprising providing to cells of the patient an antisense oligonucleotide complementary to a marker of the invention. The markers are useful for assessing if a patient is afflicted with ovarian cancer, which involves comparing the level of expression of a marker in a patient sample and a normal level of expression of the marker in a control non-ovarian cancer sample. A difference between the marker corresponds to a secreted protein or to a transcribed polynucleotide or its portion. The level of expression of the marker is assessed by detecting the presence in the sample, a protein or protein fragment corresponding to the marker. The presence of protein or protein fragment is detected using an antibody that specifically binds with the protein or protein fragment. Alternatively, the level of expression of the marker is assessed by detecting the presence of a transcribed polynucleotide which anneals with the marker or anneals with a portion of the polynucleotide comprising the marker, under stringent conditions. The marker is also used for monitoring the progression of ovarian cancer in a patient which involves detecting expression of the marker in a patient sample at a first point in time, repeating the method at a subsequent time and comparing the level of expression. The method is carried out using an ovarian tissue sample. A composition comprising a marker, polypeptide or antibody of the invention is used to treat ovarian cancer. This sequence represents a human ovarian cancer DNA marker of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 2553 BP; 690 A; 597 C; 570 G; 696 T; 0 U; 0 Other;

Query Match 66.6%; Score 595.2; DB 5; Length 2553;

Best Local Similarity 96.9%; Pred. No. 1.1e-149;

Matches 628; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

Qy 207 GGAAGAAACAAAGCCCATTTGAGTCCCTGTCAAAGAGAGAACCCCTG-AAAAA 265

Dy 695 GGAAGAAACAAAGCCCATTTGAGTCCCTGTCAAAGAGAGAACCCCTGAAAAA 636

Qy 266 CTGTTGATGTGGCAGCAGAGAAAGTGTGAAAAATTACATCTGAATACCAAGAGT 325

Dy 635 CTGTTGATGTGGCAGCAGAGAAAGTGTGAAAAATTACATCTGAATACCAAGAGT 576

Qy 326 AGAGAAATGCAGAGAGGGCTGAACGATTCAATGTACCTGTGAGCTTGAGAGTAAAG 385

Dy 575 AGAGAAATGCAGAGAGGGCTGAACGATTCAATGTACCTGTGAGCTTGAGAGTAAAG 516

Qy 386 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAAGAGTCTGTCTCTGATA 445

Dy 515 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAAGAGTCTGTCTCTGATA 456

Qy 446 ACAACCTATGTTAACTTGGTAAAGCTGAAGGAAAGAGCTCAAGATTTGGTTTGAATG 505

Dy 455 ACAACCTATGTTAACTTGGTAAAGCTGAAGGAAAGAGCTCAAGATTTGGTTTGAATG 396

Qy 506 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAAAACTGAAAAAGAGGAGCGGAT 565

Dy 395 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAAAACTGAAAAAGAGGAGCGGAT 336

Qy 566 TTGGGATTTGTACAAAGTTTCAGCTGGAACTGGAACTGGAACTGGAACTGGAACTGGAA 625

Db 335 TTGGGATTGTCACAGTTACAGTGGAACTGGAACCCAGAGGATACAGAGGCAAGAGA 276
QY 626 GGAAGAGAGAGAGCGGTTGGGATTCGCTGATGAAAGTTCCTGATATCTTCTGTTCTC 695
Db 275 GGAAGAGAGAGAGCGGTTGGGATTCGCTGATGAAAGTTCCTGATATCTTCTGTTCTC 216
QY 686 CAGTGTTCCTCATTTCT 745
Db 215 CAGTGTTCCTCATTTCT 156
QY 746 TGCTTACGTCCTGCTCGCAATCAGGAGCATGTACCCAGGTACATCCCACTGCGG 805
Db 155 TGCTTACGTCCTGCTCGCAATCAGGAGCATGTACCCAGGTACATCCCACTGCGG 96
QY 806 CAGCAGTTGACTTATGCTGTTTACGCTTTAAGGTTGTTGTTT 853
Db 95 GCAGCAGTTGACTTA-TGCTGTTTACGCTTTAAGGTTGTTGTTT 49

RESULT 13
AD122423
ID AD122423 standard; DNA; 1022 BP.
XX
AC AD122423;
XX
DT 22-APR-2004 (first entry)
XX
DE Rat liver differentially expressed cDNA seq id 233.
XX
KW hepatotropic; antiinflammatory; virucide; cytostatic; gene therapy;
KW differentially gene expression; liver; toxin; liver disorder;
KW biliary cirrhosis; x-linked adrenoleukodystrophy; Zellweger syndrome;
KW hepatorenal syndrome; hepatitis; hepatocarcinoma; metabolic response;
KW toxicological response; ss; EST; expressed sequence tag; rat.
XX
OS Rattus sp.
XX
PN US2003165854-A1.
XX
PD 04-SEP-2003.
XX
PF 05-DEC-2001; 2001US-00006285.
XX
PR 05-DEC-2000; 2000US-0251986P.
XX
PA (CUNN/) CUNNINGHAM M J.
PA (KASE/) KASER M R.
XX
PI Cunningham MJ, Kaser MR;
XX WPI; 2003-863697/80.
DR
PT New combination comprising a number of cDNAs that are differentially
PT expressed in a liver treated with a toxin, useful for diagnosing, staging
PT or treating liver disorders (e.g. biliary cirrhosis, hepatitis or
PT hepatocarcinoma).

XX
PS Claim 1; SEQ ID NO 233; 28pp; English.
XX
CC The invention describes a combination comprising a number of cDNAs that
CC are differentially expressed in a liver treated with a toxin and are
CC selected from any of the 514 cDNAs listed in the specification, or their
CC complements. The combination is useful in diagnosing, staging or treating
CC a liver disorder (e.g. biliary cirrhosis, x-linked adrenoleukodystrophy,
CC Zellweger syndrome, hepatorenal syndrome, hepatitis or hepatocarcinoma),
CC in monitoring diagnostic and therapeutic applications, in detecting
CC metabolic and toxicological responses, and in elucidating drug mechanism
CC of action. This sequence represents a cDNA differentially expressed in
CC liver tissues in response to treatment with a toxin.
XX
SQ Sequence 1022 BP; 309 A; 160 C; 229 G; 250 T; 0 U; 74 Other;

Query Match 55.7%; Score 498.4; DB 10; Length 1022;
Best Local Similarity 87.3%; Pred. No. 9.2e-124;
Matches 580; Conservative 0; Mismatches 81; Indels 3; Gaps 3;
QY 57 AAAGCTTGCAGAACTAAAGCAAGAAATGTTCTGCTCGTGGTGGAGACCAAGGGAATAA 116
Db 104 AAAGCTTGCAGAACTAAAGCAAGAAATGTTCTGCTCGGTTTAGAGACCAAGGGAATAA 163
QY 117 GCAGAGTCTTATCCACAGACT-CCAGGCATATCTTGAAGACATGCTGAGAGGAGGCA 175
Db 164 ACAAGATCTTATCAATAGGCTACAGGCATATCTTGAAGAACATGCTGAAGAAGAGCA 223
QY 176 ATGAAGAGATGTACTTGGGAGATGAAACAGAGAGAAAGAAACAAAGCCATTTGAGCTCC 235
Db 224 ATGAAGAGATGTACTTGGGAGATGAACTTGAGGAGAGAAACCAAGCCCTATAGAATGC 283
QY 236 CTGTCAAAGAGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAAAGAGTGG 295
Db 284 CTGTCAAAGAGGAAGAACCCCTGAAAAAATCTGTTGATATGGCATCAGAAAAAGAGTGG 343
QY 296 TGAAGATTTACATCTGAAATACCAACAGACTGAGAGAAATGCAGAGAGGGCTGAACGATTCA 355
Db 344 TAAAAATTTACATCTGGAATACCTCAAACTGAGAGAAATGCAGAGAGGGCTGAACGTTTCA 403
QY 356 ATGTACTCTGTAGCTTTGGAGAGTAAGAAAGCTGCTCG-GGCAGCTAGGTTTGGGATTCT 414
Db 404 ATGTGCTCTGTAAAGCTTTGGAGAGTAAAGAGCTGCTCGCGCAGCAGAGTTTGGAAATTCT 463
QY 415 TCAGTTTCCAAAGGAGTCTGTATCTGATAAACAACCTATGTTAACTTGGATAAGCTG 474
Db 464 TCAGTTTCCAAAGGAGTCTGTATCTGACACCAAGCCAATGGTTAACTTGGATAAAGTAA 523
QY 475 AAGGAAAGAGCTCAAGAGATTTGGTTGAATGTCTTCAATCTCCAGAAAGTCTGAGAT 534
Db 524 AAGGAAAGAGCTCAAGAGATTTGGTTGAATGTCTTCCATCTCTAGAAAGTCTGAGGAT 583
QY 535 GATGAGAACTGAAAAAGAGAGGAGCGATTTGGGATTTGTCAACAAGTTCCAGCTGGAAC 594
Db 584 GATGAGAACTGAAAAAGAGAGGAGATTTGGAAATGTGACAAAGTTCCAGCTGGAAC 643
QY 595 GGAACCAACAGAGGATACAGAGGCAAAAGAGAGGAAAAAGAGAGCGCTTTGGGATTGCC 654
Db 644 GGAACCAACAGAGGATACAGAGGCAAAAGAGAGGAAAAAGAGAGCGCTTTGGGATTGCA 703
QY 655 TGATGAAAAAGTTCCTG-ATACITTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCT 713
Db 704 TAATGAAAAAGTTCGTCCTTTCTGCCCACTAGTGGTTTCCATTTCTCAGATTTTCTT 763
QY 714 TGGT 717
Db 764 GGT 767

RESULT 14
AAC00738
ID AAC00738 standard; cDNA; 471 BP.
XX
AC AAC00738;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 736.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EF1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-00200610.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:19:52 ; Search time 4142.15 Seconds
(without alignments) 10458.088 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

Db	672	CTTGCCGAACTAAAGCAAGATGCTTCTGCTGCTGTTGGAGACCAAGGGAATAAGCAA	731
Qy	121	GATCTTATCCACAGACTCCAGGCATATCTTTGAAGAACATCTGAAAGAGGAGGCAATAGAA	180
Db	732	GATCTTATCCACAGACTCCAGGCATATCTTTGAAGAACATCTGAAAGAGGAGGCAATAGAA	791
Qy	181	GAAGATGCTACTGGAGATGAAACAGAGAGAGAGAAACAAGGCCATTGAGCTCCCTGTC	240
Db	792	GAAGATGTA CTGGGAGATGAACACAGAGAGAGAGAAACAAGGCCATTGAGCTCCCTGTC	851
Qy	241	AAAGAGAGAGAACCCCTGGAAGAACTGTTGATCTGGCAGCAGAGAGAGAAAGTGGTGAAA	300
Db	852	AAAGAGAGAGAACCCCTGGAAGAACTGTTGATCTGGCAGCAGAGAGAGAAAGTGGTGAAA	911
Qy	301	ATTACATCTGAAATACACAGACTGAGAGAAATGAGAGAGGGCTGAAACGATTCAATGTA	360
Db	912	ATTACATCTGAAATACACAGACTGAGAGAAATGAGAGAGGGCTGAAACGATTCAATGTA	971
Qy	361	CTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTT	420
Db	972	CTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTT	1031
Qy	421	CCACAAAGCTGCTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAA	480
Db	1032	CCACAAAGCTGCTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAA	1091
Qy	481	AGAGCTCAAGATTTGGTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG	540
Db	1092	AGAGCTCAAGATTTGGTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG	1151
Qy	541	AAACTGAAAAAGAGAGAGGAGCGATTTGGGATTTGTCAAGTTGAGTGGAACTGGAAACC	600
Db	1152	AAACTGAAAAAGAGAGAGGAGCGATTTGGGATTTGTCAAGTTGAGTGGAACTGGAAACC	1211
Qy	601	ACAGAGATACAGAGGCAAG	660
Db	1212	ACAGAGATACAGAGGCAAG	1271
Qy	661	AAAGTTCTGATATCTTCTGTTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCT	720
Db	1272	AAAGTTCTGATATCTTCTGTTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCT	1331
Qy	721	ATATATGCTTAATGCAAGTCATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	780
Db	1332	ATATATGCTTAATGCAAGTCATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1391
Qy	781	CCCAGGTACATCCATGAACCTGGCAGCAGTTTGACTTATTTGCTGTTTCAAGG	840
Db	1392	CCCAGGTACATCCATGAACCTGGCAGCAGTTTGACTTATTTGCTGTTTCAAGG	1451
Qy	841	TTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT	894
Db	1452	TTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT	1505
RESULT 2			
LOCUS HSA409089 Homo sapiens HCC-1 gene. 1557 bp DNA linear PRI 11-MAY-2001			
DEFINITION AJ409089			
ACCESSION AJ409089.1 GI:13940309			
VERSION HCC-1 gene; hepatocellular carcinoma; nuclear protein.			
KEYWORDS Homo sapiens (human)			
SOURCE Homo sapiens			
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE 1			
AUTHORS Choong, M.L., Tan, L.K., Lo, S.L., Ren, E.C., Ou, K., Ong, S.E., Liang, R.C., Seow, T.K. and Chung, M.C.			
TITLE An integrated approach in the discovery and characterization of a novel nuclear protein over-expressed in liver and pancreatic tumors			
JOURNAL FEBS Lett. 496 (2-3), 109-116 (2001)			

MEDLINE	21255661		
PUBMED	11356193		
REFERENCE	2 (bases 1 to 1557)		
AUTHORS	Choong, M.L.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-FEB-2001) Choong M.L., National University of Singapore, Bioprocessing Technology Centre, MD 11 Level 5, 10 Medical Drive, 117597, SINGAPORE		
FEATURES	Location/Qualifiers		
source	1..1557		
	/organism="Homo sapiens"		
	/mol_type="Genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="7"		
	/map="7q22"		
	/cell_line="HCC-M"		
	/cell_type="hepatocellular carcinoma"		
	/tissue_type="liver"		
gene	1..1557		
	/genes="HCC-1"		
promoter	1..687		
	/genes="HCC-1"		
	/evidence="experimental"		
misc_feature	305..594		
	/genes="HCC-1"		
	/note="non-functional upstream open reading frame"		
	/function="tightens and regulates translation of HCC-1 gene"		
repeat_region	446..453		
	/rpt_type="DIRECT"		
mRNA	661..1557		
	/genes="HCC-1"		
CDS	688..1320		
	/genes="HCC-1"		
	/codon_start=1		
	/product="HCC-1 protein"		
	/protein_id="CAC37950.1"		
	/db_xref="GI:13940310"		
	/db_xref="GOA:P82979"		
	/translation="MATETVELHLKLAELKQELARGLFTKGIKODLIHRLQAYLEE HAEENAEEDVLGDETEEBETKPIELPVKEBEPKTDVAAEKVKVITSEIPTQER MOKRAERFNPVSVLESKKAARARFGISSVPTKGLSSDNKPMVNLDKLKERARFGLN VSSISRSKSDDEKLKKRKFVITVSAGTGTEDTEAKRKAERPFIA"		
3'UTR	1321..1557		
	/genes="HCC-1"		
	/evidence="experimental"		
repeat_region	1358..1365		
	/rpt_type="DIRECT"		
polyA_site	1542		
	/genes="HCC-1"		
	/evidence="experimental"		
ORIGIN			
Query Match 100.0%; Score 894; DB 9; Length 1557;			
Best Local Similarity 100.0%; Pred. No. 9.5e-203; Indels 0; Gaps 0;			
Matches 894; Conservative 0; Mismatches 0;			
Qy	1	GGGAGTGGAGTGGAGGGGTAAAGATGGCGGACCGAGACGGTGGAGCTCCATAAGCTAAAG	60
Db	664	GGGAGTGGAGTGGAGGGGTAAAGATGGCGGACCGAGACGGTGGAGCTCCATAAGCTAAAG	723
Qy	61	CTTGCCGAACTAAAGCAAGATGCTTGTGCTGCTGTTGGAGACCAAGGGAATAAGCAA	120
Db	724	CTTGCCGAACTAAAGCAAGATGCTTGTGCTGCTGTTGGAGACCAAGGGAATAAGCAA	783
Qy	121	GATCTTATCCACAGACTCCAGGCGATCTTGAAGAACATCTGTAAGAGGAGGCAATAGAA	180
Db	784	GATCTTATCCACAGACTCCAGGCGATCTTGAAGAACATCTGTAAGAGGAGGCAATAGAA	843
Qy	181	GAAGATGTA CTGGGAGATGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	240
Db	844	GAAGATGTA CTGGGAGATGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	903

```
Qy 241 AAAGAGGAGAACCCCTGAAAAAAGCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGA 300
Db 904 AAAGAGGAGAACCCCTGAAAAAAGCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGA 963
Qy 301 ATTACATCTGAAATACACAGACTGAGAGAAATGAGAGAGGCTGAAACGATTCAATGTA 360
Db 964 ATTACATCTGAAATACACAGACTGAGAGAAATGAGAGAGGCTGAAACGATTCAATGTA 1023
Qy 361 CCTGTGAGCTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGTTTGGGATTTCTTCAGTT 420
Db 1024 CCTGTGAGCTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGTTTGGGATTTCTTCAGTT 1083
Qy 421 CCAACAAAGCTCTGTCATCTGATAACAAAGCTATGTTAACTTGGATAAGCTGAAGAA 480
Db 1084 CCAACAAAGCTCTGTCATCTGATAACAAAGCTATGTTAACTTGGATAAGCTGAAGAA 1143
Qy 481 AGAGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
Db 1144 AGAGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 1203
Qy 541 AAATGAAAAAGAGGAGGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACTGGAACC 600
Db 1204 AAATGAAAAAGAGGAGGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACTGGAACC 1263
Qy 601 ACAGAGATACAGAGGCAAGAGAGGAGAAAGAGAGAGCGCTTTGGGATTTGCCTGATGA 660
Db 1264 ACAGAGATACAGAGGCAAGAGAGGAGAAAGAGAGAGCGCTTTGGGATTTGCCTGATGA 1323
Qy 661 AAAGTCTCTGATCTTTCTGTTCTCCAGTGTTCCTTCTCTCTCTCTCTCTCTCTCTCT 720
Db 1324 AAAGTCTCTGATCTTTCTGTTCTCCAGTGTTCCTTCTCTCTCTCTCTCTCTCTCTCT 1383
Qy 721 ATATATGCTAAATGACAGCTCATGTGCTACGCTCTGCTCGCAATGAGGAGCATGTA 780
Db 1384 ATATATGCTAAATGACAGCTCATGTGCTACGCTCTGCTCGCAATGAGGAGCATGTA 1443
Qy 781 CCCAGGTACATCATGAACTGGCGCAGAGTTTGACTTATGCTTATGCTTTCAGCTTTAAG 840
Db 1444 CCCAGGTACATCATGAACTGGCGCAGAGTTTGACTTATGCTTATGCTTTCAGCTTTAAG 1503
Qy 841 TTGTTGTTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTG 894
Db 1504 TTGTTGTTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTG 1557

RESULT 3
LOCUS BD276528 1071 bp DNA linear PAT 17-JUL-2003
DEFINITION FULL-LENGTH MOLECULES EXPRESSED IN HUMAN TISSUES.
ACCESSION BD276528
VERSION BD276528.1 GI:33086296
KEYWORDS JP 2002543839-A/31.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1071)
AUTHORS Shih,L., Lu,D.A.M., Lal,P., Batra,S., Tang,T.Y.,
Yang,J., Azimzai,Y., Reddy,R., Henry, Yue and Baughn,M.R.
FULL-LENGTH MOLECULES EXPRESSED IN HUMAN TISSUES
Patent: JP 2002543839-A 31 24-DEC-2002;
INCYTE GENOMICS INC, Henry YUE, Tom Y TANG, Preeti LAL, Roopa REDDY,
Sajeew BATRA, Mariah R BAUGHN, Junning YANG, Yalda AZIMZAI, Dyoung Aina
M LU, Janice AU-YOUNG, Leo L SHIH
OS Homo sapiens
PN JP 2002543839-A/31
PD 24-DEC-2002
PF 12-MAY-2000 JP 2000618453
PR 14-MAY-1999 US 09/311937, 14-MAY-1999 US 09/311940, PR
14-MAY-1999 US 09/311894
PI leo l shih, janice l au-young, dyoung aina
m lu, preeti lal, sajeew
```

```
PI batra,
PI tom y tang, junning yang, yalda azimzai, roopa reddy, henry PI
yue, mariah r baughn
CC This description about <220> can't be interpreted CC <220>
CC <221> misc feature
CC <223> Incyte ID No: 4093555CB1
FH Key Location/Qualifiers.
FEATURES
    source
        1..1071
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
ORIGIN
Query Match 99.7%; Score 891; DB 6; Length 1071;
Best Local Similarity 100.0%; Pred. No. 4.9e-202;
Matches 891; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 AGTGGAGTGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTT 63
Db 163 AGTGGAGTGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTT 222
Qy 64 GCCGAACTAAAGCAAGAATGTCTTCTGCTGTTTGGAGACCAAGGGAATAAAGCAAGAT 123
Db 223 GCCGAACTAAAGCAAGAATGTCTTCTGCTGTTTGGAGACCAAGGGAATAAAGCAAGAT 282
Qy 124 CTATATCCAGAGCTCCAGGCATATCTTTGAAGAAATGCTGTAAGAGAGGCAAAATGAAGAA 183
Db 283 CTATATCCAGAGCTCCAGGCATATCTTTGAAGAAATGCTGTAAGAGAGGCAAAATGAAGAA 342
Qy 184 GATGTACTGGGAGATGAACAGAGGAGAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAAA 243
Db 343 GATGTACTGGGAGATGAACAGAGGAGAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAAA 402
Qy 244 GAGGAAGAAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAAGATT 303
Db 403 GAGGAAGAAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAAGATT 462
Qy 304 ACATCTGAAATACACAGACTGAGAGAAATGAGAGAGGCTGAAACGATTCATGTACCT 363
Db 463 ACATCTGAAATACACAGACTGAGAGAAATGAGAGAGGCTGAAACGATTCATGTACCT 522
Qy 364 GTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCCA 423
Db 523 GTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCCA 582
Qy 424 ACAAAGGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGCAAGA 483
Db 583 ACAAAGGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGCAAGA 642
Qy 484 GCTCAAGATTTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAA 543
Db 643 GCTCAAGATTTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAA 702
Qy 544 CTGAAAAAGAGGAGGAGCGATTTGGGATTTGTCAAGATTCAGCTGGAACCTGGAACCA 603
Db 703 CTGAAAAAGAGGAGGAGCGATTTGGGATTTGTCAAGATTCAGCTGGAACCTGGAACCA 762
Qy 604 GAGGATACAGAGGCAAGAGAGGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAGA 663
Db 763 GAGGATACAGAGGCAAGAGAGGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAGA 822
Qy 664 GTTCCTGATATTTCTGTTCTTCCAGTGTTTTCCATTTCTCTCTCTCTCTCTCTCTCT 723
Db 823 GTTCCTGATATTTCTGTTCTTCCAGTGTTTTCCATTTCTCTCTCTCTCTCTCTCTCT 882
Qy 724 TATGCTTAATGACAGCTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 783
Db 883 TATGCTTAATGACAGCTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 942
Qy 784 CAGGTACATCCATGAACTCGGGCAGAGTTTGACTTATGCTGTTTTCAGCTTTAAGGTTG 843
Db 943 CAGGTACATCCATGAACTCGGGCAGAGTTTGACTTATGCTGTTTTCAGCTTTAAGGTTG 1002
```



```
QY 338 AGAGGGCTGAACGATTCAATGTACTCTGTGAGCTGGAGAGTAAGAAGCTGCTCGGGCAG 397
Db |||||
QY 955 AGAGGGCTGAACGATTCAATGTACTCTGTGAGCTGGAGAGTAAGAAGCTGCTCGGGCAG 1014
Db |||||
QY 398 CTAGGTTTGGGATTTCTTCAGTTCCCAACAAAGGCTGTCTCATCTGATAACAAACCTATGG 457
Db |||||
QY 1015 CTAGGTTTGGGATTTCTTCAGTTCCCAACAAAGGCTGTCTCATCTGATAACAAACCTATGG 1074
QY 458 TTAACCTGGATAGCTGAAGGAAGAGCTCAAGAGATTTGGTTTGAATGCTCTTCAATCT 517
Db |||||
QY 1075 TTAACCTGGATAGCTGAAGGAAGAGCTCAAGAGATTTGGTTTGAATGCTCTTCAATCT 1134
QY 518 CCAGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGGAAGGAGCGATTTGGGATTTGCA 577
Db |||||
QY 1135 CCAGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGGAAGGAGCGATTTGGGATTTGCA 1194
QY 578 CAAGTTCAAGCTGGAACCTGGAAACACAGAGGATACAGAGGCCAAAGAGAGGAAAAAGAGCAG 637
Db |||||
QY 1195 CAAGTTCAAGCTGGAACCTGGAAACACAGAGGATACAGAGGCCAAAGAGAGGAAAAAGAGCAG 1254
QY 638 AGGCTTTGGGATGCTGTGATGAAAGTTCTCTGATATCTTCTCTCCAGTCTTTTCCA 697
Db |||||
QY 1255 AGGCTTTGGGATGCTGTGATGAAAGTTCTCTGATATCTTCTCTCCAGTCTTTTCCA 1314
QY 698 TTCTCTCTCTCTCTCTCTCTCTGTCACATATATGCTTAATGACACAGTCAATGTCCTACGTCCT 757
Db |||||
QY 1315 TTCTCTCTCTCTCTCTCTCTCTGTCACATATATGCTTAATGACACAGTCAATGTCCTACGTCCT 1374
QY 758 GCTCGCAATGAGGAGCATGTATCCCCAGGTACA 791
Db |||||
QY 1375 GCTCGCAATGAGGAGCATGTATCCCCAGGTACA 1408

RESULT 7
AC073063
LOCUS
DEFINITION Homo sapiens BAC clone RP11-136B3 from 7, complete sequence.
AC073063
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 84113)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED
9847074
REFERENCE
2 (bases 1 to 84113)
Armstrong,J., Maupin,R. and Long,N.
The sequence of Homo sapiens BAC clone RP11-136B3
Unpublished (2001)
REFERENCE
3 (bases 1 to 84113)
Waterston,R.H.
Direct Submission
Submitted (08-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 84113)
Waterston,R.H.
Direct Submission
Submitted (18-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 84113)
Waterston,R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 84113)
Waterston,R.
Direct Submission
Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 84113)
Waterston,R.
Direct Submission
Submitted (04-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 18, 2001 this sequence version replaced gi:13431102.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0136B03
-----
```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 84113)
Waterston,R.
Direct Submission
Submitted (04-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 18, 2001 this sequence version replaced gi:13431102.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0136B03

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7 . send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-114P12, 2000 bp overlap;
the clone sequenced to the left is GSI-259H13, 200 bp overlap.
Actual start of this clone is at base position 96082 of GSI-259H13;
actual end is at base position 84113 of RP11-136B3.

FEATURES

	source
1. .84113	Location/Qualifiers
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="7"	
/map="7"	
/clone="RP11-136B3"	
/clone_lib="RP11-11"	
22. .56	
/rpt_family="AT_rich"	
46. .330	
/rpt_family="Alu"	
335. .405	
/rpt_family="MER1_type"	
545. .613	
/rpt_family="MER1_type"	

	repeat_region
repeat_region	
repeat_region	
repeat_region	
repeat_region	

```
repeat_region 614. .780 /rpt_family="Alu"
repeat_region 781. .1078 /rpt_family="Alu"
repeat_region 1080. .1114 /rpt_family="MER1_type"
repeat_region 1115. .1406 /rpt_family="Alu"
repeat_region 1407. .1540 /rpt_family="MER1_type"
misc_feature 1833. .2211 /note="match to EST BE147565 (NID:g8610289)"
misc_feature 2322. .2394 /note="match to EST BE147565 (NID:g8610289)"
repeat_region 2337. .2489 /rpt_family="Alu"
repeat_region 2497. .2619 /rpt_family="MIR"
repeat_region 2620. .2644 /rpt_family="(T)n"
repeat_region 2926. .2991 /rpt_family="MIR"
repeat_region 3055. .3355 /rpt_family="Alu"
repeat_region 3374. .3395 /rpt_family="(TTTTTA)n"
repeat_region 3552. .3852 /rpt_family="Alu"
repeat_region 3853. .3972 /rpt_family="Alu"
repeat_region 4009. .4388 /rpt_family="T2_type"
repeat_region 4419. .4731 /rpt_family="Alu"
repeat_region 4786. .5085 /rpt_family="Alu"
repeat_region 5203. .5238 /rpt_family="L1"
repeat_region 5239. .5651 /rpt_family="L1"
repeat_region 5642. .5663 /rpt_family="(TAAA)n"
misc_feature 5885. .6027 /note="match to EST BE795688 (NID:g10216886)"
misc_feature 5887. .6025 /note="match to EST BG722929 (NID:g14002116)"
misc_feature 5887. .6025 /note="similar to EST BG917776 (NID:g14298252)"
misc_feature 5990. .6025 /note="match to EST BG748931 (NID:g14059584)"
repeat_region 6106. .6382 /rpt_family="L2"
misc_feature 6330. .6334 /note="match to EST BF894562 (NID:g12286021)"
repeat_region 6383. .6682 /rpt_family="Alu"
repeat_region 6683. .6744 /rpt_family="L2"
repeat_region 6745. .7046 /rpt_family="Alu"
repeat_region 7017. .7042 /rpt_family="(GAAA)n"
repeat_region 7047. .7139 /rpt_family="L2"
misc_feature 7251. .7849 /note="match to EST AV717144 (NID:g10814296)"
repeat_region 7435. .7732 /rpt_family="Alu"
repeat_region 7735. .8052 /rpt_family="Alu"
misc_feature 7907. .8001 /note="match to EST AV717144 (NID:g10814296)"
misc_feature 7927. .8032 /note="match to EST AV717144 (NID:g10814296)"

repeat_region 8017. .8086 /note="match to EST BE147565 (NID:g8610289)"
repeat_region 8118. .8234 /rpt_family="A-rich"
repeat_region 8219. .8241 /rpt_family="Alu"
repeat_region 8245. .8543 /rpt_family="AT-rich"
misc_feature 8300. .8302 /rpt_family="Alu"
repeat_region 8713. .9008 /note="match to EST AV717144 (NID:g10814296)"
repeat_region 9055. .9190 /rpt_family="Alu"
repeat_region 9191. .9501 /rpt_family="Alu"

Query Match 72.4%; Score 647.4; DB 9; Length 84113;
Best Local Similarity 90.4%; Pred. No. 6.9e-144;
Matches 782; Conservative 0; Mismatches 46; Indels 37; Gaps 7;

Qy 12 GAGGGTTAACAGATGCGACGAGACGTTGGAGCTCCATAAGCTAAAGCTTGCCGAAC 71
Db 74969 GAGGGTTAACAGATGCTTAAGATGTTGGAGCTCCATAAGCTTGCTAAACT 75028
Qy 72 AAAGCAAGATGTTCTCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCA 131
Db 75029 AAAGCAAGATGTTCTTCTGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCA 75087
Qy 132 CAGACTCCAGGCATCTTGAAGCAATGCTGAAGAGGAGGCAATCAAGCAAGATGTTACT 191
Db 75088 CAGACTCCAGGCATCTTGAAGCAATGCTGAAGAGGAGGCAATCAAGCAAGATGTTACT 75144
Qy 192 GGGAGATGAACAGAGGAGGAAGAAACAAAGCCCATTTGAGCTCCCTCTCAAGAGGAGGA 251
Db 75145 AGGAGATGAACAGAGGAGGAAGAAACAAAGCCCATTTGAGCTCCCTCTCAAGAGGAGGA 75191
Qy 252 ACCCCCTG-AAAAAATGTTGATGTGTCAGAGAGAGAAAGTGGTGAATTAATCATCTG 310
Db 75192 ACCCCCTGAAAAAATGTTGATGTGTCAGAGAGAGAAAGTGGTGAATTAATCATCTG 75251
Qy 311 AATATCCACAGCTGAGAGAAATGCAGAGAGGCTGAACGATTCATCAATGTCTGTGAGCT 370
Db 75252 AATATCCACAGGCTGAGAGAAATGCAGAGAGGCTGAACGATTCATCAATGTCTGTGAGCT 75311
Qy 371 TGGAGAGTAAGAAAGCTGCTCGGCGACGTAGTTGGGATTTCTTCAGTTCCAAACAAAG 430
Db 75312 TGGAGAGTAAGAAAGCTGCTCAGCAGCTAGTTGGGATTTCTTCAGTTCCAAACAAAG 75370
Qy 431 GTCTGTCAATCGATAACAAACCTATGTTAACTTTGGATTAAGCTGAAGGAAAGAGCTCAAA 490
Db 75371 GTCTGTCAATCG- -AACACACCTATGTTAACTTTGGATTAAGCTGAAGGAAAGAGCTCAAA 75428
Qy 491 GATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGCTGAAGATGATGAGAAACTGAAA 550
Db 75429 GATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGCTGAAGATGATGAGAAACTGAAA 75488
Qy 551 AGAGGAGGAGCGAATTTGGGATTTGTCAAGATTCAGCTGGAACCTGGAAACACAGAGATA 610
Db 75489 AGAGGAGGAGCGAATTTGGGATTTGTCAAGATTCAGCTGGAACCTGGAAACACAGAGATA 75548
Qy 611 CAGAGGCAAGAGAGGAAAGAGCAGCGCTTTGGGATTTGCTGTGATGAAAAGTTCTCTG 670
Db 75549 CAGAGGCAAGAGAGGAAAGAGCAGCGCTTTGGGATTTGCTGTGATGAAAAGTTCTCTG 75608
Qy 671 ATACTTTCTCTTCCAGTGTCTTCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 730
Db 75609 ATACTTTCTCTTCCAGTGTCTTCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 75668
Qy 731 AATGCAACATGTCGTGCTTACGTCCTCGCAATGAGGAGGAGCATGTACCCAGGTAC 790
Db 75669 AA-----TGCAGTCTTTCCTTGCATGAGGAGGAGCATGTACCCAGGTAA 75712
```


JOURNAL	Biochem. Biophys. Res. Commun. 292 (3), 593-600 (2002)
MEDLINE	21920340
PUBMED	11922608
REFERENCE	2 (bases 1 to 633)
AUTHORS	Fukuda, S., Wu, D.W., Stark, K. and Pelus, L.M.
TITLE	Direct Submission
JOURNAL	Submitted (19-FEB-2002) Microbiology/Immunology, Indiana University, School of Medicine, 1044 West Walnut Street, Indianapolis, IN 46202-5121, USA
FEATURES	Location/Qualifiers
source	1..633
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/tissue_type="primary human breast cancer"
CDS	1..633
	/notes="nuclear protein with SAP motif"
	/codon_start=1
	/product="proliferation associated cytokine-inducible protein CIP29"
	/protein_id="AA009686.1"
	/db_xref="GI:20069116"
	/translation="MATPEVELKHKLAELOKBLAQLETKGIKODLIHRLQAYLEE HAREANEEDVLGDTEETETKLELPVKEEPEPTKDVAEEKVKVKTISEI MOKRAERPNVPVLSLEKKAARAARFGISSVPTKGLSDNKPWNLDKLEKRAORFGLN VSSISRSBDEKLLKKRKERFGIVTSAGTGTEDTEAKRKRKAERFGIA"
ORIGIN	
Query Match	70.8%; Score 633; DB 9; Length 633;
Best Local Similarity	100.0%; Pred. No. 2e-140;
Matches 633; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	25 ATGCGCACGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGCAGCACTTAAAGCAAGATGT 84
Db	1 ATGCGCACGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGCAGCACTTAAAGCAAGATGT 60
Qy	85 CTGTGCTGTTGGAGACAAGGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCA 144
Db	61 CTGTGCTGTTGGAGACAAGGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCA 120
Qy	145 TATCTTCAAGAACATGCTGAAGAGGAGGCAATGAAGAAGATGTACTGGGAGATGAACA 204
Db	121 TATCTTCAAGAACATGCTGAAGAGGAGGCAATGAAGAAGATGTACTGGGAGATGAACA 180
Qy	205 GAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAAGAGGAAGAACCCCTTGAAAAA 264
Db	181 GAGGAGAGAAACAAAGCCATTGAGCTCCCTGTCAAAGAGGAGAACCCCTTGAAAAA 240
Qy	265 ACTGTTGATGTGGCAGCAGAGAAGAAAGTGGTGAATAATCATCTGAAATACCAAGACT 324
Db	241 ACTGTTGATGTGGCAGCAGAGAAGAAAGTGGTGAATAATCATCTGAAATACCAAGACT 300
Qy	325 GAGGAATGAGAGAGGCTGAACGATTCANATGACTCTGAGCTTGGAGAGTGAAGAA 384
Db	301 GAGGAAATGAGAGAGGCTGAACGATTCANATGACTCTGAGCTTGGAGAGTGAAGAA 360
Qy	385 GCTGCTCGGCAGCTAGCTTTGGGATTTCTTCAGTTCCAAACAAAGGCTGTGATCTGAT 444
Db	361 GCTGCTCGGCAGCTAGCTTTGGGATTTCTTCAGTTCCAAACAAAGGCTGTGATCTGAT 420
Qy	445 AACAAACCTATGTTTAACTTTGGATPAAGCTGAAGGAAAGAGCTCAAGATTTGGTTGAAT 504
Db	421 AACAAACCTATGTTTAACTTTGGATPAAGCTGAAGGAAAGAGCTCAAGATTTGGTTGAAT 480
Qy	505 GTCTCTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACCTGAAAAAGAGGAGGACCGA 564
Db	481 GTCTCTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACCTGAAAAAGAGGAGGACCGA 540
Qy	565 TTTGGGATTTGCACAGTTCAAGCTTGGAACTGGACACAGAGGATACAGAGGCAAGAAAG 624
Db	541 TTTGGGATTTGCACAGTTCAAGCTTGGAACTGGAACTGGAACTGGAACTGGAACTGGAA 600
Qy	625 AGGAAAAAGACAGACGCGCTTTGGGATTTGCGCTGA 657

Db 601 AGGAAAGACGAGCGCTTTGGGATTGCCTGA 633

|||||

RESULT 10
AC145981

LOCUS
DEFINITION AC145981 220280 bp DNA linear HTG 01-AUG-2003
Pan troglodytes chromosome UNK clone RP43-13G23, *** SEQUENCING IN
PROGRESS ***, 46 unordered pieces.

ACCESSION
VERSION AC145981.1 GI:33386932
KEYWORDS HTG: HTGS PHASE1.

SOURCE
ORGANISM Pan troglodytes (chimpanzee)

REFERENCE
TAXID 9606
TITLE The sequence of Pan troglodytes clone
JOURNAL Unpublished
AUTHORS Wilson, R.K.
REFERENCE 2 (bases 1 to 220280)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: C_PT013G23

* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1338: contig of 1338 bp in length
* 1339 1438: gap of unknown length
* 1439 2637: contig of 1199 bp in length
* 2638 2737: gap of unknown length
* 2738 4065: contig of 1328 bp in length
* 4066 4185: gap of unknown length
* 4186 5254: contig of 1089 bp in length
* 5255 5354: gap of unknown length
* 5355 7079: contig of 1724 bp in length
* 7079 8271: gap of unknown length
* 8271 8372: contig of 1093 bp in length
* 8372 9517: gap of unknown length
* 9517 10946: contig of 1145 bp in length
* 10946 12865: contig of 1329 bp in length
* 12865 12964: gap of unknown length
* 12965 15018: contig of 1819 bp in length
* 15018 15117: gap of unknown length
* 15118 16694: contig of 2053 bp in length
* 16694 16794: gap of unknown length
* 16794 18715: contig of 1576 bp in length
* 18715 18915: gap of unknown length
* 18915 21332: contig of 1921 bp in length
* 21332 21333: gap of unknown length
* 21333 23240: contig of 2518 bp in length
* 23240 23341: gap of unknown length
* 23341 24932: contig of 1808 bp in length
* 24932 25032: gap of unknown length
* 25032 25033: contig of 1592 bp in length
* 25033 27092: gap of unknown length
* 27092 28000: contig of 2060 bp in length

```

* 27093 27192: gap of unknown length
* 27193 28362: contig of 1170 bp in length
* 28363 28462: gap of unknown length
* 28463 29890: contig of 1428 bp in length
* 29891 31733: gap of unknown length
* 31734 31833: contig of 1743 bp in length
* 31834 34959: gap of unknown length
* 34960 35059: contig of 1126 bp in length
* 35060 37949: gap of unknown length
* 37950 38049: contig of 2890 bp in length
* 38050 41435: gap of unknown length
* 41436 41535: contig of 3386 bp in length
* 41536 46033: gap of unknown length
* 46034 46133: contig of 4498 bp in length
* 46134 49523: gap of unknown length
* 49524 49623: contig of 3390 bp in length
* 49624 54041: gap of unknown length
* 54042 54140: contig of 4417 bp in length
* 54141 58313: gap of unknown length
* 58314 58413: contig of 4173 bp in length
* 58414 62876: gap of unknown length
* 62877 62976: contig of 4463 bp in length
* 62977 69574: gap of unknown length
* 69575 75666: contig of 6598 bp in length
* 75667 75766: gap of unknown length
* 75767 81362: contig of 5992 bp in length
* 81363 81463: gap of unknown length
* 81464 85272: contig of 5596 bp in length
* 85273 91974: gap of unknown length
* 91975 92074: contig of 3810 bp in length
* 92075 96689: gap of unknown length
* 96690 103977: contig of 6602 bp in length
* 103978 104077: gap of unknown length
* 104078 109657: contig of 4615 bp in length
* 109658 109757: gap of unknown length
* 109758 116696: contig of 5580 bp in length
* 116697 116796: contig of 6939 bp in length
* 116797 124010: gap of unknown length
* 124011 124110: contig of 7214 bp in length
* 124111 132127: gap of unknown length
* 132128 132227: contig of 8017 bp in length
* 132228 139538: gap of unknown length
* 139539 139638: contig of 8017 bp in length
* 139639 148819: gap of unknown length
* 148820 148919: contig of 7111 bp in length
* 148920 157172: gap of unknown length
* 157173 157272: contig of 8253 bp in length
* 157273 168568: gap of unknown length
* 168569 168668: contig of 11296 bp in length
* 168669 179527: gap of unknown length
* 179528 179627: contig of 10859 bp in length
* 179628 189941: gap of unknown length
* 189942 189941: contig of 9314 bp in length
* 189942 189942: contig of 9314 bp in length
* 189942 201123: gap of unknown length
* 201123 201222: contig of 12081 bp in length
* 201223 220280: gap of unknown length
* 220280 220280: contig of 19058 bp in length.

FEATURES
    source
    1..220280
    /organism="Pan troglodytes"
    /mol_type="genomic DNA"
    /db_xref="taxon:9598"
    /chromosome="UNK"
    /clone="RP43-13G23"
    misc_feature 1..1338
    /note="assembly_name:Contig25"
    misc_feature 1439..2637
    /note="assembly_name:Contig33"
    misc_feature 2738..4065
    /note="assembly_name:Contig39"

```

```

misc_feature 4166..5254
/note="assembly_name:Contig40"
misc_feature 5355..7078
/note="assembly_name:Contig42"
misc_feature 7179..8271
/note="assembly_name:Contig44"
misc_feature 8372..9516
/note="assembly_name:Contig45"
misc_feature 9617..10945
/note="assembly_name:Contig46"
misc_feature 11046..12864
/note="assembly_name:Contig49"
misc_feature 12965..15017
/note="assembly_name:Contig51"
misc_feature 15118..16693
/note="assembly_name:Contig52"
misc_feature 16794..18714
/note="assembly_name:Contig53"
misc_feature 18815..21332
/note="assembly_name:Contig54"
misc_feature 21433..23240
/note="assembly_name:Contig55"
misc_feature 23341..24932
/note="assembly_name:Contig56"
misc_feature 25033..27092
/note="assembly_name:Contig57"
misc_feature 27193..28362
/note="assembly_name:Contig58"
misc_feature 28463..29890
/note="assembly_name:Contig59"
misc_feature 29991..31733
/note="assembly_name:Contig60"
misc_feature 31834..34959
/note="assembly_name:Contig61"
misc_feature 35060..37949
/note="assembly_name:Contig62"
misc_feature 38050..41435
/note="assembly_name:Contig63"
misc_feature 41536..46033
/note="assembly_name:Contig64"
misc_feature 46134..49523
/note="assembly_name:Contig65"
misc_feature 49624..54040
/note="assembly_name:Contig66"
misc_feature 54141..58313
/note="assembly_name:Contig67"
misc_feature 58414..62876
/note="assembly_name:Contig68"
misc_feature 62977..69574
/note="assembly_name:Contig69"
misc_feature 69675..75666
/note="assembly_name:Contig70"
misc_feature 75767..81362
/note="assembly_name:Contig71"
misc_feature 81463..85272
/note="assembly_name:Contig72"
misc_feature 85373..91974
/note="assembly_name:Contig73"
misc_feature 92075..96689
/note="assembly_name:Contig74"

```

```

Query Match      69.6%; Score 622.4; DB 2; Length 220280;
Best Local Similarity 88.5%; Pred.No. 6.4e-138;
Matches 768; Conservative 0; Mismatches 61; Indels 39; Gaps 7;

QY 12 GAGGGGTAACAGATGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGCAGAACT 71
    |||||
Db 147627 GAGGGGTAACAGATGCGTAACGTGAGATGGTGGAGCTCCGTAAGCTTGCCTAACT 147686
    |||||

QY 72 AAAGCAAGATGCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCA 131
    |||||
Db 147687 AAAGCAAGATGCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCA 147746
    |||||

```

```
Qy 132 CAGACTCCAGCATATCTTGAAGAACATGCTGAGAGGAGGCAATGAAGAGATGTACT 191
Db 147747 CAGACTCCAGCATATCTTGAAGAACATGCTGAGAGGAGGCAAT 147803
Qy 192 GGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGA 251
Db 147804 AGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGA 147850
Qy 252 ACCCCCTG-AAAAATCTGTTGATGTGGCAGCAGAGAAAGTGGTGAATTCATCTG 310
Db 147851 ACCCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAAAGTGGTGAATTCATCTG 147910
Qy 311 AATATACACACATGAGAGATGAGAGAGGCTGACCAATTCATGTTACCTGTGAGCT 370
Db 147911 AATATACACAGGCTGAGAGATGAGAGAGGCGCAATTCAGTGTACCTGTGAGCT 147970
Qy 371 TGGAGAGTAAG- --AAAGTCTCGGCGAGCTAGGTTGGATTCTTCAGTTCACCAAA 427
Db 147971 TGGAGAGTAAGTCAAGTAAATCAGGAGCTAGGTTGGATTCTTCAGGCAACAA 148030
Qy 428 AAGTCTGTCTATGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAGAGCTC 487
Db 148031 AA-CTCTGTCTATG- --AACACACCTATGGTTAACTTGGATAAGCGAAGGAAGAGCTC 148087
Qy 488 AAGATTTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGA 547
Db 148088 AAGATCTGGTTTGAATGTGTCTTCAATCTCCGAAAGTCTGAAGATGATAAGAACTGA 148147
Qy 548 AAGAGAGGAGGAGCGATTGGGATTGTCAAGTCTCAGCTGGAACCTGGAACACACAGG 607
Db 148148 AAGAGAGGAGGAGCGATTGGGATTGTCAAGTCTCAGCTGGAACCTGGAACCCACAGG 148207
Qy 608 ATACAGAGGCAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTGCTGATGAAAAGTTC 667
Db 148208 ATACAGAGGCAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTGCTGATGAAAAGTTC 148267
Qy 668 CTGATCTTTCTGTTTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 727
Db 148268 TTGATCTTTCTGTTTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 148327
Qy 728 CCTAAATGCACAGTCATGTGCTACGTCCTGCGAATGAGGAGGAGCATGTACCCAGG 787
Db 148328 CTAAA-----TGACGTCCTTGCCTTGAATGAGGAGGAGCATGTACCCAGG 148371
Qy 788 TACATCCATGAATCGCGCAGCAGTTGACTTATGCTTTTTCAGCTTTAAGGTTGTTGT 847
Db 148372 TAAACTGTGAATCGCGCAGCAGTTGACTTATGCTTTTTCAGCTTTAAGGTTGTTGT 148431
Qy 848 GTTTTGTGTTTGAATGTTGCTTGT 875
Db 148432 GTTTTGTGTTTGAATGTTGCTTGT 148459

RESULT 11
AC134830 155666 bp DNA linear ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-498G16 from chromosome 19, complete
sequence.
ACCESSION AC134830
VERSION AC134830.2 GI:34495140
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
VanBrunt,A.
REFERENCE 1 (bases 1 to 155666)
TITLE The sequence of Mus musculus BAC clone RP24-498G16
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 155666)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
```

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

3 (bases 1 to 155666)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (01-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 155666)
Wilson,R.K.
Direct Submission
Submitted (07-SEP-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

5 (bases 1 to 155666)
Wilson,R.
Direct Submission
Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 7, 2003 this sequence version replaced gi:23396352.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watson.wustl.edu
----- Summary Statistics

Center project name: M_BB0498G16

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

FEATURES
source
1..155666
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="19"
/map="19"
/clone="RP24-498G16"
/clone_lib="RPCI-24"
1..118
/rpt_family="MaLR"
136..410
/rpt_family="MaLR"
472..670
/rpt_family="ERV1"
1983..2092
/rpt_family="MER2_type"
2103..2222
/rpt_family="BC1_MM"
2242..2546
/rpt_family="MER2_type"

repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

repeat_region	4917.4984	/rpt_family="L1"
repeat_region	5489.5558	/rpt_family="MaLR"
repeat_region	5559.5986	/rpt_family="MaLR"
repeat_region	5887.6640	/rpt_family="MaLR"
repeat_region	6679.6727	/rpt_family="MaLR"
repeat_region	6985.7169	/rpt_family="B2"
repeat_region	9274.9616	/rpt_family="MaLR"
repeat_region	9949.10089	/rpt_family="B4"
repeat_region	11286.11443	/rpt_family="ERV1"
repeat_region	11691.12093	/rpt_family="MaLR"
repeat_region	12146.12311	/rpt_family="L1"
repeat_region	12312.12494	/rpt_family="ERVK"
repeat_region	13916.14409	/rpt_family="B2"
repeat_region	15237.15387	/rpt_family="MIR"
repeat_region	16072.16182	/rpt_family="MaLR"
repeat_region	16183.16651	/rpt_family="RMER62B"
repeat_region	16652.16842	/rpt_family="MaLR"
repeat_region	17041.17109	/rpt_family="ERV1"
repeat_region	17202.17280	/rpt_family="B2"
repeat_region	17439.17674	/rpt_family="L1"
repeat_region	19313.19449	/rpt_family="Alu"
repeat_region	20160.20251	/rpt_family="ID"
repeat_region	20693.20835	/rpt_family="B4"
repeat_region	20849.21036	/rpt_family="B2"
repeat_region	21284.21740	/rpt_family="L1"
repeat_region	21741.22138	/rpt_family="MaLR"
repeat_region	22139.22242	/rpt_family="L1"
repeat_region	22243.22421	/rpt_family="B2"
repeat_region	22422.24053	/rpt_family="L1"
repeat_region	24110.24269	/rpt_family="L1"
repeat_region	24546.24762	/rpt_family="L1"
repeat_region	25008.25215	/rpt_family="MaLR"
repeat_region	26954.27056	/rpt_family="ERVU"
repeat_region	28629.29006	/rpt_family="MaLR"
repeat_region	29259.29379	

repeat_region	/rpt_family="MaLR"	30433..30712	
repeat_region	/rpt_family="B4"	31744..32053	
repeat_region	/rpt_family="L1"	32097..32379	
repeat_region	/rpt_family="MaLR"	32380..32695	
repeat_region	/rpt_family="MaLR"	33964..34098	
repeat_region	/rpt_family="Alu"	34526..34700	
repeat_region	/rpt_family="B4"	35974..37364	
repeat_region	/rpt_family="L1"	37436..37859	
repeat_region	/rpt_family="B2"	37779..37851	
repeat_region	/rpt_family="B4"	38383..39139	
misc_feature	/note="CpG island (%GC=72.4, o/e=0.96, #CpGs=99)"		
repeat_region		41242..41562	
repeat_region	/rpt_family="MaLR"	41672..41741	
repeat_region	/rpt_family="MER1_type"	43225..43814	
repeat_region	/rpt_family="L1"	44602..45138	
repeat_region	/rpt_family="MaLR"	45382..45479	
repeat_region	/rpt_family="Alu"	47954..48356	
repeat_region	/rpt_family="MaLR"	48596..48779	
repeat_region	/rpt_family="L1"	50526..50554	
Query Match			
Best Local Similarity 68.0%; Score 607.6; DB 10; Length 155666;			
Matches 743; Conservative 0; Mismatches 144; Indels 19; Gaps 3;			
Qy	5	GTGGAGTGGGGGTACAAAGATGGCGGACCGAGACGGTGAGAGCTCCATAAGCTTAAAGCTTG	64
Db	27075	GTGGACTGGAGCGGCATCAAGATGGCGGCGGAGACGGTGAGAGCTCCAAAGCTGAAGCTTG	27134
Qy	65	CCGAACCTAAAGCAAGAATCTCTTGCTCGTGGTTTGAGACCAAGGGAATAAGCAAGATC	124
Db	27135	CTGAACCTAAGCAAGAATCTCTTGCTCGTGGTTTAGAGACCAAGGGAATAAAACAAGATC	27194
Qy	125	TTATCCACAGACTCCAGGCATATCTTTGAAGAACATGCTGTGAAGAGAGGCAATGAAGAAG	184
Db	27195	TTATAAATAGGCTGCAGGCGCTATCTTTGAAGACCATTCGAGAACGAAGAAGCAATGAAGAAG	27254
Qy	185	ATGTACTGGGAGATGAACAGAGGAGAGAACCAAGCCCATTCAGCTCCCTGTCAAG	244
Db	27255	ATGTACTGGGAGATGAACCTGAGGAAGAAGAACCAAGCCCATTCAGCTCCCTGTCAAG	27314
Qy	245	AGGAAGAACCCCTCGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTA	304
Db	27315	AGGAAGAACCTCTGAAAGGCTGTTGATATGTCATCGGAAAGAGGTTGGTAAAGATTA	27374
Qy	305	CATCTGAATATACACAGACTGAGAGATGTCAGAGAGGCTGAACGATTCATGTACCTG	364
Db	27375	CATCTGGAATACCTCAGACTGAAGAAATGAGAGAGGCTGAACGATTCATGTACCTG	27434
Qy	365	TGAGCTTGGAGATGAAGAAGCTGCTCGGCGAGCTAGGTTTCGGGATTTCTTCAGTTC	424
Db	27435	TAGACCTGGAGATGAAGAAGCTGCCCGGCGGCTAGGTTTCGATTTCTTCAGTTC	27494
Qy	425	CAAAAGGCTGTCTCATCTGATAACAAACCTATGTTAACTTGATAGCTGAAGGAAGAG	484
Db	27495	CAAAAGGTTTATCATCTGACCAAGGCTATGGTTAACTGATGAATAACTAAAGGAAGAG	27554


```
repeat_region 5743..5839 /rpt_family="L1"
repeat_region 5924..6010 /rpt_family="RMR1B"
repeat_region 6496..6633 /rpt_family="RMR1B"
repeat_region 6779..6884 /rpt_family="MIR"
repeat_region 6885..7304 /rpt_family="RMR17C"
repeat_region 7305..7333 /rpt_family="MIR"
repeat_region 9080..9263 /rpt_family="L1"
repeat_region 9503..9905 /rpt_family="MaLR"
repeat_region 12380..12477 /rpt_family="Alu"
repeat_region 12721..13257 /rpt_family="MaLR"
repeat_region 14045..14634 /rpt_family="L1"
repeat_region 16118..16187 /rpt_family="MER1_type"
repeat_region 16297..16617 /rpt_family="MaLR"
misc_feature 18720..19476 /notes="CpG island (GC=72.4, o/e=0.96, #CpGs=99)"
repeat_region 20081..20114 /rpt_family="B4"
repeat_region 20300..20423 /rpt_family="Alu"
repeat_region 20495..21885 /rpt_family="B2"
repeat_region 22610..22761 /rpt_family="L1"
repeat_region 23159..23333 /rpt_family="B4"
repeat_region 23761..23895 /rpt_family="Alu"
repeat_region 25164..25479 /rpt_family="MaLR"
repeat_region 25575..25762 /rpt_family="MaLR"
repeat_region 25806..26115 /rpt_family="L1"
repeat_region 27147..27426 /rpt_family="B4"
repeat_region 28480..28600 /rpt_family="MaLR"
repeat_region 28853..29230 /rpt_family="MaLR"
repeat_region 30803..30905 /rpt_family="ERV1"
repeat_region 32644..32851 /rpt_family="MaLR"
repeat_region 33097..33313 /rpt_family="L1"
repeat_region 33590..33749 /rpt_family="L1"
repeat_region 33806..35437 /rpt_family="L1"
repeat_region 35438..35616 /rpt_family="B2"
repeat_region 35617..35720 /rpt_family="L1"
repeat_region 35721..36118 /rpt_family="MaLR"
repeat_region 36119..36575 /rpt_family="L1"
repeat_region 36823..37010 /rpt_family="L1"

/rpt_family="B2"
37021..37166 /rpt_family="B4"
37608..37699 /rpt_family="ID"
38410..38547 /rpt_family="Alu"
40185..40420 /rpt_family="L1"
40579..40657 /rpt_family="B2"
40693..40766 /rpt_family="ERV1"
40965..41155 /rpt_family="MaLR"
41156..41624

Query Match 68.0%; Score 607.6; DB 10; Length 159681;
Best Local Similarity 82.0%; Pred. No. 2.2e-134;
Matches 743; Conservative 0; Mismatches 144; Indels 19; Gaps 3;

QY 5 GTGGAGTGAAGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATTAAGCTTAAGCTTG 64
Db 30784 GTGGAGTGAAGGGTAAACAAGATGGCGCGCGAGACGGTGGAGCTCCACAAGCTGAAGCTTG 30725
QY 65 CCGAAGTAAAGCAAGAAATGCTTGTCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATC 124
Db 30724 CTGAAGTAAAGCAAGAAATGCTTGTCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATC 30665
QY 125 TTATCCACAGACTCCAGGCATATCTTGAAGAACATCTGTAAGAGAGGAGGCAAAATGAAGAAG 184
Db 30664 TTATAAATAGGCTGAGGCTATCTTGAAGACCATCGAAGAGAGCAAAATGAAGAAG 30605
QY 185 ATGTACTGGGAGATGAACAGAGGAAGAAAGAAACAAAGGCCATTGAGCTCCCTGTCAAAG 244
Db 30604 ATGTACTGGGAGATGAACAGAGGAAGAAAGAAACAAAGGCCATGAACTGCCTGTAAAG 30545
QY 245 AGGAAGAACCCCTGAAAAAATGTTGATGTGGCAGCAGAGAGAAAGTGGTGAAGATTA 304
Db 30544 AGGAAGAACCCCTGAAAAAGGCTGTTGATGTGGCAGCAGAGAGGTTGGTGAAGATTA 30485
QY 305 CATCTGAAATACCAAGAGCTGAGAGAAATGCAGAGAGGGCTGAACGATTCAATGTACCTG 364
Db 30484 CATCTGGAATACCTCAGCTGAAAGAAATGCAGAGAGGGCTGAACGATTCAATGTACCTG 30425
QY 365 TGAGCTTGGAGAGTAAAGAAAGCTGCTCGGCGACAGTGGTGGGATTTCTTCAGTTCCAA 424
Db 30424 TAAGCCTGGAGAGTAAAGAAAGCTGCTCGGCGCGGCTAGGTTTGGAAATTTCTTCAGTTCCAA 30365
QY 425 CAAAAGTCTGCTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAG 484
Db 30364 CAAAAGGTTTATCATCTGACCAACGCTATGTTAACTTGGATAAGCTGAAGGAAAGAG 30305
QY 485 CTCAAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAAC 544
Db 30304 CACAGAGATTTGGTTTGAATGTCTCTTCAATCTCTGAGAAAGTCTGAGGATGATGAGAGC 30245
QY 545 TGAAGAGAGGAAAGGAGCGATTTGGGATTTGTCAAGAGTTCAGCTGGAACCTGGAACCCACAG 604
Db 30244 TGAAGAAACCGAAGGAGAGATTTGGGATTTGTGCAAGTTCAGCTGGAACCTGGAACCCACAG 30185
QY 605 AGGATACAGAGGCAAGAGAGAGAAAGAGCAGAGCGCTTGGGATTTGGGATTTGGTGAAGAG 664
Db 30184 AGGATACAGAGGCAAGAGAGAGAAAGAGAGCAGAGCGCTTGGGATTTGGGATTTGGTGAAGAG 30125
QY 665 TTCCTGATATCTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 724
Db 30124 TTCGTG---CTTTCTGCACCATAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 30068
QY 725 ATGCCTAAATGCACAGTCATGTGCCTACGTC-----CTGCCTCGCAATGA 769
Db 30067 ATACATACACACCCCTACCTACGTAATCTATCTATACATACACAGTCATGTGCCTCACAGTGA 30008
```


Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INL at: <http://image.llnl.gov>
 Series: IRAK Plate: 67 Row: e Column: 6
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13384729.

FEATURES

Location/Qualifiers
 1..908
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="MGC:41057 IMAGE:1382954"
 /tissue_type="Mammary gland, lactating mouse"
 /clone_lib="Soares_mammary_gland_NLMG"
 /lab_host="DH10B"
 /notes="vector: pT7T3-Pac"
 1..908
 /gene="1110005A23Rik"
 /db_xref="LocusID:66118"
 /db_xref="MGI:1913368"
 5..637
 /gene="1110005A23Rik"
 /codon_start=1
 /product="RIKEN cDNA 1110005A23"
 /protein_id="AAH27510.1"
 /db_xref="GI:20380578"
 /db_xref="LocusID:66118"
 /db_xref="MGI:1913368"
 /translation="MAAEVELHLKLAEILKQELARGLETGKIKQDLINRLQAYLED
 HQEAEENFVLDGTEEEPKPIELPVEEPEPKAVDMSEKKVVKITSGTPQTER
 MOKRAEFNVPSLESKLAARAFGISVPTKGLSDTKPMVNLDKLKERAQRFGLN
 VSSISRSKSEDEKLKKRERFGIVTSSAGTGTTEAKKRKRAERFGIA"

gene

CDS

ORIGIN

Query Match 66.4%; Score 593.2; DB 10; Length 908;
 Best Local Similarity 81.8%; Pred. No. 6.2e-131;
 Matches 728; Conservative 0; Mismatches 143; Indels 19; Gaps 3;
 QY 21 CAAGATGGCAGCAGACGGTGGAGCTCCATAGCTAAAGCTGCCGAATCAAGCAAGA 80
 DB 1 CAAGATGGCGCCGAGACGGTGGAGCTCCAAAGCTGAAGCTTGCTGAATCAAGCAAGA 60
 QY 81 ATGTCTTGCTGCTGCTGGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGACTCCA 140
 DB 61 ATGTCTTGCTGCTGCTGGTTAGACCAAGGGAATAAACAAGATCTTATTAATAGGCTGCA 120
 QY 141 GGCAATCTTTGAAGAA CATGCTGAAGAGAGGCGCAATGAAGAGATGTTACTGGGAGATGA 200
 DB 121 GGCCTATCTTGAAGACCATGCAAGAAAGAAAGCAAAATGAAGAGATGTTACTGGGAGATGA 180
 QY 201 AACAGGAGAGAGAAACAAAGCCCATGAGCTCCCTGTCAAGAGGAAGAACCCCTGTA 260
 DB 181 AACTGAGGAAGAAACAAAGCCCATAGAACTGCCCTGTTAAAGAGGAAGAACCTCTGTA 240
 QY 261 AAAAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATACATCTGAATACACACA 320
 DB 241 AAAGCTGTTGATATGCAATCGGAAGAGAGTGGTGAAGATACATCTGGATACCTCA 300
 QY 321 GACTGAGAGATGCAAGAGGCGGTGAACGAATCAATGTACCTGTGAGCTGGAGAGTAA 380
 DB 301 GACTGAAGAATGCAAGAGGCGGTGAACGAATCAATGTACCTGTAAGCCCTGGAGAGTAA 360
 QY 381 GAAAGCTGCTCGGCAGCTAGTGTGGATTTCTTCAGTTCACAAAGCTCTGTCATC 440
 DB 361 GAAGGCTGCGCGCGCGCTAGGTTTGGAAATTTCTCAGTTCCAAACAAAGGTTTATCATC 420
 QY 441 TGATAACAAACCTATGTTTAACTTTGGATAAGCTCAAGAAAGAGCTCAAGAGATTTGGTTT 500
 DB 421 TGACACCAAGCCTATGTTTAACTTGGATTAACCTGAAGAAAGAGCA CAGAGATTTGGTTT 480
 QY 501 GAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAATCTGAAAAAGGAGGA 560
 DB 481 GAATGTCTTCTCCATCTCTAGAAAGTCTGAGGATGATGAGAAGCTGAAGAAACGGAAGGA 540

QY 561 GCGATTTCGGATTGTCCAAAGTTTCAGCTGGAACTGGAACACACAGAGGATACAGAGCAAA 620
 DB 541 GAGATTTCGGATTGTCCAAAGTTTCAGCTGGAACTGGAACACACAGAGGATACAGAGCAAA 600
 QY 621 GAAGAGAAAAAGAGCAGAGCGCTTTGGGATTCGCTGATGAAAAAGTTCTCTGATACATCTCTG 680
 DB 601 GAAAAAGAAAAAGCAGAGCGCTTTGGAAATTCGATAATGAAAAAGTTCTGCTG---CTTTCTG 657
 QY 681 TTCTCCAGTGTTCCTCCATTTCTCTCTCTCTCTCTCTCTCATATATATGCGCTAAATGACACAG 740
 DB 658 CACCATAGTGTTCCTCCATTTCTCAAAATTTTCTTGGTTTATATATACATACACACCCCT 717
 QY 741 TCATGTGCGCTACGTC-----CTGCTCGCAATCAGGAGCAGCATGTACCCCA 785
 DB 718 ACCTACTACCTACTCTACATACACAGTCATGTCCTCACAGTGGAGAGCATGTACGTTT 777
 QY 786 GGTACATCATGAACCTGCGGACGACGTTTGACTTATTCGTGTTTACGTTTAAAGTTGTT 845
 DB 778 GGTACATCCTTGAATTCAGTAGCAGTTTATTTAACTGCTCTTCTGCGTTTAAAGATTGTT 837
 QY 846 GT-GTTTTTGTTCGATTATGTTGCTTGTATATAAAAAAATAGAAA 894
 DB 838 GTGGTTTTTGTTCGTTTAAATTTATTTTCTGCTTATTAATAAAAAAGGAGAAAA 887
 RESULT 15
 AC022200/c 148152 bp DNA linear HTG 09-MAY-2001
 LOCUS Homo sapiens chromosome 8 clone RP11-262017 map 8, WORKING DRAFT
 DEFINITION SEQUENCE, 4 unordered pieces.
 AC022200 GI:13518207
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 148152)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 8, clone RP11-262017
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 148152)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bada, F., Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A., Choepe, I., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArillano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marguis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Olivari, T. M., Peterson, K., Pierre, N., Pisani, C., Pollata, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 TITLE Direct Submission
 JOURNAL Submitted (26-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Apr 3, 2001 this sequence version replaced gi:13488035.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIER
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information

 Center project name: L5196

Page Blank (uspio)


```
Db 6158 AAATACACAGGCTGAGAGATCGAGAGAGGCGCGAACAATTCAGTGTACCTGTGAGCT 6099
QY 371 TGAGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAACAAAAG 430
Db 6098 TGAGAGTAAGAAAGCTGCTCAGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCAACAAA- 6040
QY 431 GTCTGTCTCATGATACAAACCTATGGTTAACTTGGATAGCTGAAGGAAGGCTCAAA 490
Db 6039 GTCTGTCTCATG--AACACACCTATGGTTAACTTGGATAGCGAAGGAAGGCTCAAA 5982
QY 491 GATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAA 550
Db 5981 GATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAA 5922
QY 551 AGAGGAAGGAGCGATTGGGATTTGTCAAGTTTCAGCTGGAACTGGAACCCACAGAGATA 610
Db 5921 AGAGGAAGGAGCGATTGGGATTTGTCAAGTTTCAGCTGGAACTGGAACCCACAGAGATA 5862
QY 611 CAGAGCAAGAGAGAGAAAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTTCTCTG 670
Db 5861 CAGAGCAAGAGAGAGAAAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTTCTCTG 5802
QY 671 ATACTTCTCTGCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 730
Db 5801 ATACTTCTCTGCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5742
QY 731 AAATGACACAGTATGTGCTTACGTCCTGCTCGCAATGAGGGAGCATGTACCCAGGTAC 790
Db 5741 AA-----TGCAGTCTTGCCTTGCATGAGGGAGCAGGTACCCAGGTAA 5698
QY 791 ATCCATGAATCGCGCAGCAGTTTGACTTATGCTGTTTTCAGCTTTAAAGTTTGTGTT 850
Db 5697 AACTGTGAATCGCGCGCAGTTTGACTTATGCTGTTTTCAGCTTTAAAGTTTGTGTT 5638
QY 851 TTTGTTTTTGAATTATGTTGTTGTT 875
Db 5637 TTTGTTTTTGAATTATGTTGTTGTT 5613

RESULT 2
US-09-949-016-12964/c
; Sequence 12964, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 12964
; LENGTH: 22303
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12964

Query Match 72.4%; Score 647.4; DB 4; Length 22303;
Best Local Similarity 90.4%; Pred. No. 6.3e-186;
Matches 782; Conservative 0; Mismatches 46; Indels 37; Gaps 7;

QY 12 GAGGGGTAAACAAGATGCGACCGGAGCGGTGGAGCTCCATAAGCTAAAGCTTCCGGA 71
Db 6441 GAGGGGTAAACAAGATGCGTAACTGAGATGGTGGAGCTCCATAAGCTGGAAGCTTGTAA 6382
```

```
QY 72 AAAGCAAGAAATCTCTTGTCTCGTGGTTTGGAGACCAGGGAATAAAAGCAAGATCTTATCCA 131
Db 6381 AAAGCAAGAAATCTCTTGTCTCGTGGTTTGGAGACCAGGGAATAAAAGCAAGATCTTATCCA 6323
QY 132 CAGACTCCAGGCGATATCTTTGAAGAAACATGCTCAAGAGGAGGCAAAATGAAAGAGATGACT 191
Db 6322 CAGACTCCAGGCGATATCTTTGAAGAAACATGCTCAAGAGGAGGCAAAAT---GAAGATGACT 6266
QY 192 GGGAGATGAAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTCTGTCAAGAGGAAGA 251
Db 6265 AGGAGATGAAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTCTGTCAAGAGGAAGA 6219
QY 252 ACCCCCTG-AAAGAACTGTGTGTCGACAGAGAAAGTGGTGAAGAAATTCATCTG 310
Db 6218 ACCCCCTGAAAGAACTGTGTGTCGACAGAGAAAGTGGTGAAGAAATTCATCTG 6159
QY 311 AAATACACAGACTGAGAGAAATGCAGAAAGAGGCTGAAACGATTTCAATGTACCTGTGAGCT 370
Db 6158 AAATACACAGACTGAGAGAAATGCAGAAAGAGGCGCGAACAATTCAGTGTACCTGTGAGCT 6099
QY 371 TGAGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAACAAAAG 430
Db 6098 TGAGAGTAAGAAAGCTGCTCAGGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCAACAAA- 6040
QY 431 GTCTGTCTCATGATACAAACCTATGTTTAACTTTGGATAAGCTGAAAGGAAGAGCTCAAA 490
Db 6039 GTCTGTCTCATG--AACACACCTATGTTTAACTTTGGATAAGCGAAGGAAGAGCTCAAA 5982
QY 491 GATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAA 550
Db 5981 GATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAA 5922
QY 551 AGAGGAAGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACCCACAGAGATA 610
Db 5921 AGAGGAAGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACCCACAGAGATA 5862
QY 611 CAGAGCAAGAGAGAGAAAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTTCTCTG 670
Db 5861 CAGAGCAAGAGAGAGAAAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTTCTCTG 5802
QY 671 ATACTTCTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCATATATGCCT 730
Db 5801 ATACTTCTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCATATATATA 5742
QY 731 AAATGACACAGTCAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 790
Db 5741 AA-----TGCAGTCTTGCCTTGCATGAGGGAGCAGGTACCCAGGTAA 5698
QY 791 ATCCATGAATCGCGCAGCAGTTTGACTTATGCTGTTTTCAGCTTTAAAGTTTGTGTT 850
Db 5697 AACTGTGAATCGCGCGCAGTTTGACTTATGCTGTTTTCAGCTTTAAAGTTTGTGTT 5638
QY 851 TTTGTTTTTGAATTATGTTGTTGTT 875
Db 5637 TTTGTTTTTGAATTATGTTGTTGTT 5613

RESULT 3
US-09-621-976-3627
; Sequence 3627, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3627
; LENGTH: 553
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36...452
US-09-621-976-3627

Query Match      58.9%; Score 527; DB 4; Length 553;
Best Local Similarity 99.8%; Pred. No. 3.2e-150;
Matches 538; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy  4 AGTGGAGTGAAGGCTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTT 63
Db  15 AGTGGAGTGAAGGCTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTT 74
Qy  64 GCCGAACTAAAGCAAGAAATGCTTGTCTCGTGGTTTGGAGACCAAGGGAATAAACAAGAT 123
Db  75 GCCGAACTAAAGCAAGAAATGCTTGTCTCGTGGTTTGGAGACCAAGGGAATAAACAAGAT 134
Qy  124 CTTATCCACAGACTCCAGGCAATATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAA 183
Db  135 CTTATCCACAGACTCCAGGCAATATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAA 194
Qy  184 GATGTACTGGAGATGAAACAGAGCAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 243
Db  195 GATGTACTGGAGATGAAACAGAGCAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 254
Qy  244 GAGGAAGAACCCCTGAAAAAATCTTGTGATGTGCGCAGAGAAAGAGTGGTGAATAAT 303
Db  255 GAGGAAGAACCCCTGAAAAAATCTTGTGATGTGCGCAGAGAAAGAGTGGTGAATAAT 314
Qy  304 ACATCTGAATACACAGACTGAGAGAAATGCGAAGAGGCGTGAACGATCAATGTACCT 363
Db  315 ACATCTGAATACACAGACTGAGAGAAATGCGAAGAGGCGTGAACGATCAATGTACCT 374
Qy  364 GTGAGCTTGGAGATGAAGAGCTGCTCGGCACGTA- GGTTCGGGATTTCTTCAGTTCC 422
Db  375 GTGAGCTTGGAGATGAAGAGCTGCTCGGCACGTAAGCTTGGGATTTCTTCAGTTCC 434
Qy  423 AACAAAAAGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAG 482
Db  435 AACAAAGGCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAG 494
Qy  483 AGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATAGA 541
Db  495 AGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATAGA 553

RESULT 4
US-09-513-999C-736
; Sequence 736, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 736
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25...471
; OTHER INFORMATION:
; US-09-621-976-3627
```

```
; NAME/KEY: misc_feature
; LOCATION: 354
; FEATURE:
; OTHER INFORMATION: Y=c or t
US-09-513-999C-736

Query Match      52.3%; Score 467.6; DB 4; Length 471;
Best Local Similarity 99.8%; Pred. No. 3.7e-132;
Matches 467; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy  4 AGTGGAGTGAAGGCTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTT 63
Db  4 AGTGGAGTGAAGGCTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTT 63
Qy  64 GCCGAACTAAAGCAAGAAATGCTTGTCTCGTGGTTTGGAGACCAAGGGAATAAACAAGAT 123
Db  64 GCCGAACTAAAGCAAGAAATGCTTGTCTCGTGGTTTGGAGACCAAGGGAATAAACAAGAT 123
Qy  124 CTTATCCACAGACTCCAGGCAATATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAA 183
Db  124 CTTATCCACAGACTCCAGGCAATATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAA 183
Qy  184 GATGTACTGGAGATGAAACAGAGCAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 243
Db  184 GATGTACTGGAGATGAAACAGAGCAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 243
Qy  244 GAGGAAGAACCCCTGAAAAAATCTTGTGATGTGCGCAGAGAAAGTGGTGAATAAT 303
Db  244 GAGGAAGAACCCCTGAAAAAATCTTGTGATGTGCGCAGAGAAAGTGGTGAATAAT 303
Qy  304 ACATCTGAATACACAGACTGAGAGAAATGCGAAGAGGCGTGAACGATCAATGTACCT 363
Db  304 ACATCTGAATACACAGACTGAGAGAAATGCGAAGAGGCGTGAACGATCAATGTACCT 363
Qy  364 GTGAGCTTGGAGATGAAGAAAGCTGCTCGGCACGTAAGCTTGGGATTTCTTCAGTTCCA 423
Db  364 GTGAGCTTGGAGATGAAGAAAGCTGCTCGGCACGTAAGCTTGGGATTTCTTCAGTTCCA 423
Qy  424 ACAAAGGCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAG 471
Db  424 ACAAAGGCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAG 471

RESULT 5
US-09-621-976-18639
; Sequence 18639, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18639
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 126..127
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-18639

Query Match      41.9%; Score 374.2; DB 4; Length 405;
Best Local Similarity 95.0%; Pred. No. 1e-103;
Matches 383; Conservative 11; Mismatches 7; Indels 2; Gaps 1;

Qy  7 GGAGTGAAGGCTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTCCC 66
Db  1 GGAGTGAAGGCTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTCCC 60
```



```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; CLONE: 232-463-14
; US-08-232-463-14

Query Match          9.0%; Score 80.6; DB 1; Length 7218;
Best Local Similarity 6.9%; Pred.No.1.9e-13;
Matches 29; Conservative 238; Mismatches 152; Indels 0; Gaps 0;

Qy 1 GGGAGTGAGTGGAGGTAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTAAAG 60
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 61 CTTCGGCACTAAAGCAAGATGTTCTGCTGCTGTTGGAGACCAAGGGAAATGAAGCA 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 121 GATCTATCCACAGACTCCAGGCATATCTTGAAGACATGCTGAAGAGGCGCAATCAA 180
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1319 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1260
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 181 GAAGATGTACTGGGAGTGAACAGAGCAAGAGAAACAAAGCCCATTTGAGCTCCCTGTC 240
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1259 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1200
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 241 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAA 300
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1199 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1140
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 301 ATTACATCTGAATACACAGACTGAGAGATGCAAGAGAGGCTGAACGATTCAATGTA 360
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1139 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1080
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 361 CTGTGAGCTTGGAGTAAGAAAGCTGCTCGGCAGCTAGTGTGGGATTTCTTCACT 419
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1079 RRRRRRRRRRRATGCGAGCTCCCTCGACTGCGAGCAAGCTCGGAATTAATTCGT 1021
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-949-016-35088/c
; Sequence 35088, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35088
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-35088

Query Match          8.4%; Score 75; DB 4; Length 601;
Best Local Similarity 87.6%; Pred.No.2.1e-12;
Matches 106; Conservative 0; Mismatches 10; Indels 5; Gaps 2;

Qy 12 GAGGGTTAACAAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAAGCTTGCCTGAACT 71
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 415 GAGGGTTAACAAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAAGCTTGCCTGAACT 356
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 72 AAAGCAA----GAATGCTTCTGCTGCTGTTGGAGACCAAGGGAAATAAAGCAAGATCTTA 127
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 355 AAAGTAAGAATGAATGCTTCTGCTGCTGTTGGAGACCAAGGGAAATAAAGCAAGATCTTA 297
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 128 T 128
Db |
Qy 296 T 296
Db |

RESULT 10
US-09-949-016-42456/c
; Sequence 42456, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 42456
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-42456

Query Match          8.4%; Score 75; DB 4; Length 601;
Best Local Similarity 87.6%; Pred.No.2.1e-12;
Matches 106; Conservative 0; Mismatches 10; Indels 5; Gaps 2;

Qy 12 GAGGGTTAACAAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAAGCTTGCCTGAACT 71
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 415 GAGGGTTAACAAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAAGCTTGCCTGAACT 356
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 72 AAAGCAA----GAATGCTTCTGCTGCTGTTGGAGACCAAGGGAAATAAAGCAAGATCTTA 127
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 355 AAAGTAAGAATGAATGCTTCTGCTGCTGTTGGAGACCAAGGGAAATAAAGCAAGATCTTA 297
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 128 T 128
Db |
Qy 296 T 296
Db |

RESULT 11
US-09-949-016-35089/c
; Sequence 35089, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:20:27 ; Search time 3569.42 Seconds
(without alignments)
9533.602 Million cell updates/sec

Title: US-09-788-476A-1
Perfect score: 894
Sequence: 1 99gagtggtgaggggttaa.....taataaaaaaaatagaaaa 894

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	861	96.3	871	5	BM916484
2	843.6	94.4	874	5	BM915580
3	843.6	94.4	1000	7	CO580484
4	839.8	93.9	877	5	BU849740
5	839	93.8	910	3	AF161434
6	836	93.5	844	3	CO610192
7	831.8	93.0	963	7	CO648068
8	828	92.6	1030	4	BM559381
9	826.6	92.5	895	5	BU174287
10	821	91.8	1070	4	BM460786
11	820.2	91.7	928	5	BU855435
12	817.6	91.5	843	5	BU456776
13	811.4	90.8	817	5	BU456775
14	801.4	89.6	910	5	BU157949
15	800.8	89.6	823	5	BU599301
16	800.4	89.5	914	5	BU508603
17	791.2	88.5	1006	5	BQ068156
18	784.2	87.7	842	3	BQ692014
19	784	87.7	784	3	CR593273
20	783.8	87.7	937	5	BU856660
21	782.2	87.5	883	5	BU146608
22	782	87.5	794	7	CN261593
23	780.8	87.3	795	7	CN261585
24	780.6	87.3	806	4	BG574651

25	774.6	86.6	1105	7	CK231375
26	768.6	86.0	779	5	BX344918
27	765.2	85.6	788	4	BM462456
28	762	85.2	914	5	BU902964
29	759	84.9	783	4	BG910161
30	758.8	84.9	1104	5	BM912983
31	758.4	84.8	876	5	BU855417
32	757.2	84.7	785	5	BU632453
33	757	84.7	818	5	BQ227687
34	756.8	84.7	902	5	BQ962276
35	756.6	84.6	783	6	CA775512
36	755.8	84.5	760	4	BM722636
37	752	84.1	767	5	BQ008624
38	750.8	84.0	876	5	BU194990
39	749.4	83.8	783	4	BG533012
40	749.2	83.8	791	6	CB956735
41	748.2	83.7	753	5	BM979030
42	742.2	83.0	878	5	BQ213428
43	741.4	82.9	887	4	BI255433
44	741	82.9	757	5	BU538304
45	740.4	82.8	744	7	CN479544

ALIGNMENTS

RESULT 1
BM916484
LOCUS BM916484 871 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6641858 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5482902
5', mRNA sequence.
ACCESSION BM916484
VERSION BM916484.1 GI:19366863
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 871)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2009 row: 1 column: 07
High quality sequence stop: 748.
Location/Qualifiers
1..871
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5482902"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

FEATURES

source
1..871
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5482902"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 96.3%; Score 861; DB 5; Length 871;

```
Best Local Similarity 99.3%; Pred. No. 1.1e-208;
Matches 864; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 25 ATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGCAGAACTAAAGCAAGATGT 84
    |||||
Db 1 ATGCGACCGAGACGGTGGAGCTCCATAAGCTTGCAGAACTAAAGCAAGATGT 60

Qy 85 CTTGCTCGTGGTTTGGAGACCAAGGGAATTAAGCAAGATCTTATCCACAGACTCCAGGCA 144
    |||||
Db 61 CTTGCTCGTGGTTTGGAGACCAAGGGAATTAAGCAAGATCTTATCCACAGACTCCAGGCA 120

Qy 145 TATCTTGAAGAAATGCTGAAGAGGAGGCAAAATGAAGCAAGATGCTGAGAGATGAACA 204
    |||||
Db 121 TATCTTGAAGAAATGCTGAAGAGGAGGCAAAATGAAGCAAGATGCTGAGAGATGAACA 180

Qy 205 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 264
    |||||
Db 181 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 240

Qy 265 ACTGTTGATGTCGACGACGAGAGAAAGTGGTGAATAATTACATCTGAAATACCAAGACT 324
    |||||
Db 241 ACTGTTGATGTCGACGACGAGAGAAAGTGGTGAATAATTACATCTGAAATACCAAGACT 300

Qy 325 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 384
    |||||
Db 301 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 360

Qy 385 GCTGCTCGGCGACGACTAGGTTTGGGATTTCTTCCAGTTCCAAAGAGGCTGTGTCATCTGAT 444
    |||||
Db 361 GCTGCTCGGCGACGACTAGGTTTGGGATTTCTTCCAGTTCCAAAGAGGCTGTGTCATCTGAT 420

Qy 445 AACAAACCTATGTTAACTTGGATAAGCTGAAGAAAGAGCTCAAAGATTTGGTTTGAAT 504
    |||||
Db 421 AACAAACCTATGTTAACTTGGATAAGCTGAAGAAAGAGCTCAAAGATTTGGTTTGAAT 480

Qy 505 GTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGMAAAAGAGGAAGGCGA 564
    |||||
Db 481 GTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGMAAAAGAGGAAGGCGA 540

Qy 565 TTTGGGATTTGTCACAAAGTTTCAAGCTGGAACTGGAAACCAAGAGGATACAGAGCAAGAAAG 624
    |||||
Db 541 TTTGGGATTTGTCACAAAGTTTCAAGCTGGAACTGGAAACCAAGAGGATACAGAGCAAGAAAG 600

Qy 625 AGGAAAGAGACAGACGCTTTGGATTTCCCTGTGATGAAAGTTCCTGTATCTTCTGTTCT 684
    |||||
Db 601 AGGAAAGAGACAGACGCTTTGGATTTCCCTGTGATGAAAGTTCCTGTATCTTCTGTTCT 660

Qy 685 CCAGTGTCTTCCATTTCTCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGAT 744
    |||||
Db 661 CCAGTGTCTTCCATTTCTCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGAT 720

Qy 745 GTGCTCAGTCTGCTGCAATGAGGAGCATGTACCCAGGTPACATCATGAACTGCG 804
    |||||
Db 721 GTGCTCAGTCTGCTGCAATGAGGAGCATGTACCCAGGTPACATCATGAACTGCG 780

Qy 805 GCACAGTTTACATTTATGCTTTTCAAGTTTGAAGTTTGGTGTGTTTCTTTTGTATTA 864
    |||||
Db 781 GCACAGTTTACATTTATGCTTTTCAAGTTTGAAGTTTGGTGTGTTTCTTTTGTATTA 840

Qy 865 TGTGCTTGTGTTTAAATAAAAAAATAGAAAA 894
    |||||
Db 841 TGTGCTTGTGTTTAAATAAAAAAATAGAAAA 870
```

```
RESULT 2
BM915580
LOCUS BM915580
DEFINITION AGENCOURT 6701683 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5481778
5', mRNA sequence.
ACCESSION BM915580
VERSION BM915580.1 GI:19365959
KEYWORDS EST.
SOURCE Homo sapiens (human)
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 874)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTP/DTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2006 row: m column: 11
High quality sequence stop: 668.

FEATURES
    source
        1..874
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5481778"
            /tissue_type="amelanotic melanoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_41"
            /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Library constructed by Ling Hong in the
            laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."
```

```
ORIGIN
Query Match 94.4%; Score 843.6; DB 5; Length 874;
Best Local Similarity 98.7%; Pred. No. 3.2e-204;
Matches 861; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

Qy 25 ATGGCCACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTTCGCCGAACTAAAGCAAGATGT 84
    |||||
Db 1 ATGGCCACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTTCGCCGAACTAAAGCAAGATGT 60

Qy 85 CTTGCTCGTGGTTTGGAGACCAAGGGAATTAAGCAAGATCTTATCCACAGACTCCAGGCA 144
    |||||
Db 61 CTTGCTCGTGGTTTGGAGACCAAGGGAATTAAGCAAGATCTTATCCACAGACTCCAGGCA 120

Qy 145 TATCTTGAAGAAATGCTGAAGAGGAGGCAAAATGAAGCAAGATGCTGAGAGATGAACA 204
    |||||
Db 121 TATCTTGAAGAAATGCTGAAGAGGAGGCAAAATGAAGCAAGATGCTGAGAGATGAACA 180

Qy 205 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 264
    |||||
Db 181 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 240

Qy 265 ACTGTTGATGTCGACGACGAGAGAAAGTGGTGAATAATTACATCTGAAATACCAAGACT 324
    |||||
Db 241 ACTGTTGATGTCGACGACGAGAGAAAGTGGTGAATAATTACATCTGAAATACCAAGACT 300

Qy 325 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 384
    |||||
Db 301 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 360

Qy 385 GCTGCTCGGCGACGACTAGGTTTGGGATTTCTTCCAGTTCCAAAGAGGCTGTGTCATCTGAT 444
    |||||
Db 361 GCTGCTCGGCGACGACTAGGTTTGGGATTTCTTCCAGTTCCAAAGAGGCTGTGTCATCTGAT 420

Qy 445 AACAAACCTATGTTAACTTGGATAAGCTGAAGAAAGAGCTCAAAGATTTGGTTTGAAT 504
    |||||
Db 421 AACAAACCTATGTTAACTTGGATAAGCTGAAGAAAGAGCTCAAAGATTTGGTTTGAAT 480
```

```
QY 505 GTCTCTCAATCCAGAAAGTCTGAAGATGATGAGAACTGAAGAGAGAGCGA 564
|
|
|
Db 481 GTCTCTCAATCCAGAAAGTCTGAAGATGATGAGAACTGAAGAGAGAGCGA 540
|
|
|
QY 565 TTTGGGATTTGTCACAAAGTTTCAGCTGGAACCCACAGAGGATACAGAGGCGAAAGAG 624
|
|
|
Db 541 TTTGGGATTTGTCACAAAGTTTCAGCTGGAACCCACAGAGGATACAGAGGCGAAAGAG 600
|
|
|
QY 625 AGGAAAAGACAGAGCGCTTTGGGATTCGCTGATGAAAAGTTTCCTGATACATTTCTGTCT 684
|
|
|
Db 601 AGGAAAAGACAGAGCGCTTTGGGATTCGCTGATGAAAAGTTTCCTGATACATTTCTGTCT 660
|
|
|
QY 685 CCAAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 744
|
|
|
Db 661 CCAAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
|
|
|
QY 745 GTGCCCTACGCTCTCGCTCGCAATGAGGAGCATGACCCAGGTACATCCATGAACTGCG 804
|
|
|
Db 721 GTGCCCTACGCTCTCGCTCGCAATGAGGAGCATGACCCAGGTACATCCATGAACTGCG 780
|
|
|
QY 805 GCAGCAGTTTGACTTATTTGCTGTTTTCAGCTTTT--AAGGTTGTTGTTGTTTGTGTTT 862
|
|
|
Db 781 GCAGCAGTTTGACTTATTTGCTGTTTTCAGCTTTTAAAGGTTGTTGTTGTTTGTGTTT 840
|
|
|
QY 863 TATGTTGCTGTTTAAATAAAAAAATAGAAA 894
|
|
|
Db 841 TATGTTGCTGTTTAAATAAAAAAATAGAAA 872
|
|
|
```

RESULT 3

```
CO580484
LOCUS
DEFINITION
  ILLUMIGEN MCQ 48614 Katze MMTE Macaca mulatta cDNA clone
  IBUW:18168 5' similar to Bases 522 to 900 highly similar to human
  CIP29 (Hs.410597), mRNA sequence.
```

ACCESSION

```
CO580484
```

VERSION

```
EST.
```

KEYWORDS

```
Macaca mulatta (rhesus monkey)
```

ORGANISM

```
Macaca mulatta
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
```

```
Cercopitheciinae; Macaca.
```

```
1. (bases 1 to 1000)
```

```
Katze, M.G., Thomas, M., Korth, M., Iadonato, S.P. and Magness, C.L.
```

```
Large-scale Rhesus Macaque cDNA Sequencing
```

```
Unpublished (2003)
```

```
Contact: C. Magness
```

```
Illumigen Biosciences Inc.
```

```
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
```

```
Tel: 2063780400
```

```
Fax: 2063780408
```

```
Email: cmagness@illumigen.com
```

```
Sequenced on 2004.07.02. 775 Q20 bases. Library Preparation: Prof.
```

```
Michael Katze Lab at University of Washington DNA Sequencing:
```

```
Illumigen Biosciences Inc. For further information, see
```

```
http://www.macaque.org
```

```
PCR Primers
```

```
FORWARD: CCTCACTAAAGGGAACAAA
```

```
BACKWARD: CACTATAGGCGAATGGGTA
```

```
Insert Length: 1000 Std Error: 0.00
```

```
Plate: CL000408 row: B column: 10
```

```
Seq primer: CCTCACTAAAGGGAACAAA
```

```
POLYA=Yes.
```

```
Location/Qualifiers
```

```
1..1000
```

```
/organism="Macaca mulatta"
```

```
/mol_type="mRNA"
```

```
/strain="Indian"
```

```
/db_xref="taxon:9544"
```

```
/clone="IBUW:18168"
```

```
/sex="male"
```

```
/tissue_type="testis"
```

FEATURES

```
source
```

```
/dev_stage="juvenile"
```

```
/lab_host="Electromax DH10B"
```

```
/clone_lib="Katze MMTE"
```

```
/note="Vector: pDONR 222; Site 1: BsrG I; Site 2: BsrG I;
```

```
Created from CloneMiner cDNA Library Construction kit
```

```
(catalog #18249-029)"
```

ORIGIN

```
Query Match      94.4%; Score 843.6; DB 7; Length 1000;
Best Local Similarity 97.8%; Pred. No. 3.3e-204;
Matches 855; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 21 CAAGATGCGACGACGAGCGTGGAGCTCCATAAGCTAAAGCTTCCGCAACTAAAGCAAGA 80
|
|
|
Db 5 CAAGATGCGACGACGAGCGTGGAGCTCCATAAGCTAAAGCTTCCGCAACTAAAGCAAGA 64
|
|
|
QY 81 ATGCTCTTCTCTGCTGTTTGGAGACCAAGGGGAATAAAGCAAGATCTTATCCACAGACTCCA 140
|
|
|
Db 65 ATGCTCTTCTCTGCTGTTTGGAGACCAAGGGGAATAAAGCAAGATCTTATCCACAGACTCCA 124
|
|
|
QY 141 GGCATATCTTGAAGAACATGCTGAAGAGGAGCGCAAAATGAAGAAGATGCTGGGAGATGA 200
|
|
|
Db 125 GGCATATCTTGAAGAACATGCTGAAGAGGAGCGCAAAATGAAGAAGATGCTGGGAGATGA 184
|
|
|
QY 201 AACAGAGGAAGAACAAACAAAGCCCATTGAGCTCCCTGTCAAAGAGGAAGAACCCCTTGA 260
|
|
|
Db 185 AACAGAGGAAGAACAAACAAAGCCCATTGAGCTCCCTGTCAAAGAGGAAGAACCCCTTGA 244
|
|
|
QY 261 AAAAAGCTTGTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTTACATCTGAAATACCACA 320
|
|
|
Db 245 AAAAAGCTTGTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTTACATCTGAAATACCACA 304
|
|
|
QY 321 GACTGAGAGAAATCAGAGAGGCGGTGAACGATTCATATGTACCTGTGAGCTTTGGAGAGTAA 380
|
|
|
Db 305 GACTGAGAGAAATCAGAGAGGCGGTGAACGATTCATATGTACCTGTGAGCTTTGGAGAGTAA 364
|
|
|
QY 381 GAAAGCTGCTCGGCAGCTAGGTTTGGATTTCTTCAGTTCCCAACAAAGATCTGTCTATC 440
|
|
|
Db 365 GAAAGCTGCTCGGCAGCTAGGTTTGGATTTCTTCAGTTCCCAACAAAGATCTGTCTATC 424
|
|
|
QY 441 TGATAACAAACCTATGTTTAACTTTGGAATAAGCTGAAGGAAAGAGCTCAAGATTTGGTTT 500
|
|
|
Db 425 AGATAACAAACCTATGTTTAACTTTGGAATAAGCTGAAGGAAAGAGCTCAAGATTTGGTTT 484
|
|
|
QY 501 GAATGTCTCTTCAATCTCCAGAAAGTCTGAAGAGATGATGAGAACTGAAAAAGAGGAAGA 560
|
|
|
Db 485 GAATGTCTCTTCAATCTCCAGAAAGTCTGAAGAGATGATGAGAACTGAAAAAGAGGAAGA 544
|
|
|
QY 561 GCGATTTGGGATTTGTCAAGTTCCAGCTGGAACCTGGAACCAACAGAGGATACAGAGGCAAA 620
|
|
|
Db 545 GCGATTTGGGATTTGTCAAGTTCCAGCTGGAACCTGGAACCAACAGAGGATACAGAGGCAAA 604
|
|
|
QY 621 GAAGGGAAGAGACAGAGCGCTTTGGATTTGCGCTGATGAAAGTTCTCTGATATCTTCTG 680
|
|
|
Db 605 GAAGGGAAGAGACAGAGCGCTTTGGATTTGCGCTGATGAAAGTTCTTGTATCTATCTG 664
|
|
|
QY 681 TTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTTGGTCCACATATATGCTAAATGCAAG 740
|
|
|
Db 665 TTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTTGGTCCACATATATGCTAAATGCAAG 724
|
|
|
QY 741 TCATGTGCTTACGTCCTGCGCTCGCAATGAGGAGGAGCATGACCCAGGTACATCCATGAAC 800
|
|
|
Db 725 TCACGTGCTTACGTCCTGCGCTCGCAATGAGGAGGAGCATGACCCAGGTACATCCGTTGAAC 784
|
|
|
QY 801 TGGCGCAGCAGTTTGACTTATTTGCTTTGAGCTTTAAGGTTGTTGTTTGTGTTTTTGT 860
|
|
|
Db 785 TGGCGCAGCAGTTTGACTTATTTGCTTTTCACTGTTTCACTTTAAGGTTGTTGTTTGT 844
|
|
|
QY 861 ATTATGTTGCTTGTGTTTAAATAAAAAAATAGAAA 894
|
|
|
Db 845 ATTATGTTGCTTGTGTTTAAATAAAAAAATAGAAA 878
|
|
|
```

```
RESULT 4
BU849740
LOCUS
DEFINITION AGNCOURT_10440969 NIH_MGC_109 Homo sapiens cDNA clone
IMAGE:6598412 5', mRNA sequence.
ACCESSION BU849740
VERSION BU849740.1 GI:24034703
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 877)
NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M2831 row: c column: 20
High quality sequence stop: 659.
Location/Qualifiers
1..877
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6598412"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 109"
/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

```
ORIGIN
Query Match 93.9%; Score 839.8; DB 5; Length 877;
Best Local Similarity 98.6%; Pred. No. 3e-203;
Matches 855; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY 28 GCGACCGAGCGGTGGAGCTCCATAGCTTAAGCTTGCAGACTAAAGCAAGATGCTT 87
DB 1 GCGACCGAGCGGTGGAGCTCCATAGCTTAAGCTTGCAGACTAAAGCAAGATGCTT 60
QY 88 GCTCGTGTGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCATAT 147
DB 61 GCTCGTGTGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCATAT 120
QY 148 CTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAAGATGTACTGGGAGATGAAGACAGAG 207
DB 121 CTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAAGATGTACTGGGAGATGAAGACAGAG 180
QY 208 GAAGNAGAACAAAGCCCATTTGAGCTCCCTGTCAAGAGGAGAACCCCTGAAAAAACT 267
DB 181 GAAGNAGAACAAAGCCCATTTGAGCTCCCTGTCAAGAGGAGAACCCCTGAAAAAACT 240
QY 268 GTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTACATCTGAAATACCAACAGACTGAG 327
DB 241 GTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTACATCTGAAATACCAACAGACTGAG 300
QY 328 AGAATGCAGAGAGGGCTGAACGATTCATGTAATGTACTGTGAGCTTGGAGAGTAAGAAAGCT 387
DB 301 AGAATGCAGAGAGGGCTGAACGATTCATGTAATGTACTGTGAGCTTGGAGAGTAAGAAAGCT 360
```

```
388 GCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTTCCAAACAAAAGGCTCTGTCTCATCTGATAAC 447
361 GCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTTCCAAACAAAAGGCTCTGTCTCATCTGATAAC 420
448 AAACCTATGTTAACTTTGGAATAAGCTGAAGGAAAGAGCTCAAAAGATTTGGTTTGAATGTC 507
421 AAACCTATGTTAACTTTGGAATAAGCTGAAGGAAAGAGCTCAAAAGATTTGGTTTGAATGTC 480
508 TCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGAGGAGCGATTT 567
481 TCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGAGGAGCGATTT 540
568 GGGATTTGTCACAAAGTTCAGCTGGAACTGGACACACAGAGATACAGAGGAAAGAGAGG 627
541 GGGATTTGTCACAAAGTTCAGCTGGAACTGGACACACAGAGATACAGAGGAAAGAGAGG 600
628 AAAAGAGCAGAGCGCTTTGGGATTTGCCGTGATGAAAAAGTTCTCTGATATCTTTCTTCTCCA 687
601 AAAAGAGCAGAGCGCTTTGGGATTTGCCGTGATGAAAAAGTTCTCTGATATCTTTCTTCTCCA 660
688 GTGTTTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 747
661 GTGTTTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
748 CCTACGTCCTCGCAATGAGGAGCATGTACCCAGGTACATCCATCAACTGCCGCA 807
721 CCTACGTCCTCGCAATGAGGAGCATGTACCCAGGTACATCCATCAACTGCCGCA 780
808 GCAGTTTGACTTATTTGCTGTTTTCAGCTTTAAGGTTGTTGTTTGTGTTTGTGTTTGTGTTTGT 867
781 GCAGTTTGACTTATTTGCTGTTTTCAGCTTTAAGG-TGGTGGGTTTGTGTTTGTGTTTGTGTTTGT 839
868 TGCTTGTGTTAATAAAAAAATAGAAA 894
840 TGCTTGGTATATAAAAAAATAGAAA 866
```

```
RESULT 5
AF161434
LOCUS
DEFINITION Homo sapiens HSPC316 mRNA, partial cds.
ACCESSION AF161434
VERSION AF161434.1 GI:6841281
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 910)
AUTHORS Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
Fan, H.Y., Mo, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
TITLE Human partial CDS from cd34+ stem cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 910)
AUTHORS Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
Fan, H.Y., Mo, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1999) Shanghai Institute of Hematology, Shanghai
Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,
Shanghai 200025, P. R. China
FEATURES
source
1..910
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CBLAIH07"
/cell_type="cd34+ stem cell"
/tissue_type="blood"
<1..769
/codon_start=2
/product="HSPC316"
/protein_id="AAF28994.1"
```

/db_xref="GI:6841282"
/translation="SVGRGNMATAITVELHKLKLAELQCKLARGLETGKIKDILHR
LOAVLEHAEENEDVIGDTEETETKTELPKREPEPKTVDAASKVKVITS
EIQPTMRQKRAERNFVVISLSKKAARAARVGISVVPKRSVLLINLMITWISKE
ELKDLGVSSISRSKSEDEKLUKKERKFGVITTSAGTGTTFTEARRGKEQSAUGLP
DEKFLISLVQFPFELSPFLVTPMKTCVLMCLRPAQ"

ORIGIN	Query Match	93.8%;	Score 839;	DB 3;	Length 910;
	Best Local Similarity	99.0%;	Pred. No. 4.8e-203;		
	Matches 886;	Conservative 0;	Mismatches 5;	Indels 4;	Gaps 4;
Qy	3	GAGTGGAGTGAGGGGTAAACAAGATGGCGACCGGACGCGTGGAGCTCCATAAGCTAAAGCT	62		
Db	1	GAGTGGAGTGAGGGGTAAACAAGATGGCGACCGGACGCGTGGAGCTCCATAAGCTAAAGCT	60		
Qy	63	TGCCGAACTAAAGCAAGAATGCTCTGCTGCTGGTGTGGAGACCAAGGCAATAAAGCAAGA	122		
Db	61	TGCCGAACTAAAGCAAGAATGCTCTGCTGCTGGTGTGGAGACCAAGGCAATAAAGCAAGA	120		
Qy	123	TCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGCAATGAAGA	182		
Db	121	TCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGCAATGAAGA	180		
Qy	183	AGATGTACTGGGAGATGAACAGAGGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAA	242		
Db	181	AGATGTACTGGGAGATGAACAGAGGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAA	240		
Qy	243	AGAGGAAGAACCCCTCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAANAAT	302		
Db	241	AGAGGAAGAACCCCTCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAANAAT	300		
Qy	303	TACATCTGAATAACCAAGACTGAGAGAAATGCAGAGAGGGCTGAACGATTCATGTPACC	362		
Db	301	TACATCTGAATAACCAAGACTGAGAGAAATGCAGAGAGGGCTGAACGATTCATGTPACC	360		
Qy	363	TGTGAGCTTCGGAGTAAGAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTT-C	421		
Db	361	TGTGAGCTTCGGAGTAAGAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTTCC	420		
Qy	422	CAACAAAGCTCTGTCTAT-CTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAA	480		
Db	421	CAACAAAGCTCTGTCTATCTGTATACAAACCTATGGTTAACTTGGATAAGCTTGAAGGA	480		
Qy	481	AGAGCTCAAGAATTT-GGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGA	539		
Db	481	AGAGCTCAAGAATTTGGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGA	540		
Qy	540	GAACTGAAAAGAGGAGGAGCGAATTTGGGATTTGTCACAGTTTCAGCTCGAACTGGAAC	599		
Db	541	GAACTGAAAAGAGGAGGAGCGAATTTGGGATTTGTCACAGTTTCAGCTCGAACTGGAAC	600		
Qy	600	CACAGAGCATACAGAGGCAAAAGAGAGCAAAAGAGCAGAGCGCTTTTGGGATTCGCTGATG	659		
Db	601	CACAGAGCATACAGAGGCG-AGAAGAGCAAAAGAGCAGAGCGCTTTTGGGATTCGCTGATG	659		
Qy	660	AAAAGTTCCTGATATTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTGGTCA	719		
Db	660	AAAAGTTCCTGATATTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTGGTCA	719		
Qy	720	CATATATGCTTAATGACAGTCAATGTGCCTACGTCTCTGCTCGCATGAGGAGCATGT	779		
Db	720	CATATATGCTTAAATGACAGTCAATGTGCCTACGTCTCTGCTCGCATGAGGAGCATGT	779		
Qy	780	ACCCAGGTATACATCCATGAACCTGGCGCAGAGTTTGATCTTATTTGCTGTTTCAAG	839		
Db	780	ACCCAGGTATACATCCATGAACCTGGCGCAGAGTTTGATCTTATTTGCTGTTTCAAG	839		
Qy	840	GTTGTTGTTGTTTGTGTTTTCATPATGTTGCTTGTTAATAAAAAAATAGAAA	894		
Db	840	GTTGTTGTTGTTTGTGTTTTCATPATGTTGCTTGTTAATAAAAAAATAGAAA	894		

```

RESULT 6
CR610192      844 bp      mRNA      linear      HTC 21-JUL-2004
LOCUS      full-length cDNA clone CS0CAP004Yall1 of Thymus of Homo sapiens
DEFINITION      (human).
ACCESSION      CR610192
VERSION      CR610192.1 GI:50490999
KEYWORDS      HTC; CNSLT_CDNA.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 844)
              Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
              Full-length cDNA libraries and normalization
              Unpublished
              Contact : Feng Liang Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
              Paraday Avenue
              2 (bases 1 to 844)
              Genoscope.
              Direct Submission
              Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
              - Web : www.genoscope.cns.fr)
COMMENT      1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
              end enriched, double-strand cDNA was digested with Not I and cloned
              into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
              was normalized. Library was constructed by Life Technologies, a
              division of Invitrogen.
FEATURES             Location/Qualifiers
     source           1..844
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP004Yall1"
                     /tissue_type="Thymus"
                     /plasmid="pCMVSPORT_6"
ORIGIN
Query Match      93.5%;      Score 836;      DB 3;      Length 844;
Best Local Similarity 100.0%;      Pred. No. 2.7e-202;
Matches 836;      Conservative 0;      Mismatches 0;      Indels 0;      Gaps 0;

Qy      59      AGCTTGCCGAACATAAGCAAGATGCTTGCTCGTGTGTTGGAGACCAAGGGAATAAAGC 118
         |||||
Db      1      AGCTTGCCGAACATAAGCAAGATGCTTGCTCGTGTGTTGGAGACCAAGGGAATAAAGC 60

Qy      119     AAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAATGCTGAAGAGAGGCAATG 178
         |||||
Db      61     AAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAATGCTGAAGAGAGGCAATG 120

Qy      179     AAGAAGATGCTCGGAGATGAACACAGAGGAGAGAGAAACAAAGCCATTGAGCTCCCTG 238
         |||||
Db      121     AAGAAGATGCTCGGAGATGAACACAGAGGAGAGAGAAACAAAGCCATTGAGCTCCCTG 180

Qy      239     TCAAGAGGAAGAAACCCCTGAAAAAATGTTGATGTGCAGCAGAGAGAAAGTGGTCA 298
         |||||
Db      181     TCAAGAGGAGAAACCCCTGAAAAAATGTTGATGTGCAGCAGAGAGAAAGTGGTCA 240

Qy      299     AAATTACATCTGAAATACCAAGACTGAGAGAAATCAGAAGAGGGCTGAACGATTCATG 358
         |||||
Db      241     AAATTACATCTGAAATACCAAGACTGAGAGAAATCAGAAGAGGGCTGAACGATTCATG 300

Qy      359     TACCTGTGAGCTTGGAGAGTAAAGAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAG 418
         |||||
Db      301     TACCTGTGAGCTTGGAGAGTAAAGAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAG 360

Qy      419     TTCACACAAAGGTCGTGATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGG 478
         |||||
Db      361     TTCACACAAAGGTCGTGATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGG 420

Qy      479     AAAGAGCTCAAGATTTGGTTTGAATGTCTCTTCATCTCCAGAAAGTCTGAAGATGATG 538
         |||||

```

```
|||||
421 AAAGAGCTCAAGATTTGGTTTGAATGTCCTTCAATCTCAGAAAGCTCTGAAGATGATG 480
Qy 539 AGAACTGAAAGAGAGAGGAGCGATTTGGATTGTACAAAGTTTCAGCTGGAACTGGAA 598
Db 481 AGAACTGAAAGAGAGAGGAGCGATTTGGATTGTACAAAGTTTCAGCTGGAACTGGAA 540
Qy 599 CCACAGAGGATACAGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658
Db 541 CCACAGAGGATACAGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy 659 GAAAGTTCCTGATACATTTCTGTTCTCCAGTGTTTCCATTCTCTCTCTCTCTCTCTCT 718
Db 601 GAAAGTTCCTGATACATTTCTGTTCTCCAGTGTTTCCATTCTCTCTCTCTCTCTCTCT 660
Qy 719 ACATATATGCTAAATGCACAGTCATGTGCTACAGTCTCTGCTCGCAATGAGGAGCATG 778
Db 661 ACATATATGCTAAATGCACAGTCATGTGCTACAGTCTCTGCTCGCAATGAGGAGCATG 720
Qy 779 TACCCAGGTACATCCATGAACTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 838
Db 721 TACCCAGGTACATCCATGAACTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy 839 GGTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 894
Db 781 GGTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 836
```

```
RESULT 7
CO648068
LOCUS
DEFINITION
ILLUMIGEN MQO 41334 Katze MRPB2 Macaca mulatta cDNA clone
IBUW:25134 5' similar to Bases 527 to 886 highly similar to human
CIP29 (Hs.410597), mRNA sequence.
CO648068
CO648068.1 GI:50569562
EST.
Macaca mulatta (rhesus monkey)
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Macaca.
1 (bases 1 to 963)
Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
Large-scale Rhesus Macaca cDNA Sequencing
Unpublished (2003)
Contact: C. Magness
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagness@illumigen.com
Sequenced on 2004.06.02. 750 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCCTCACTAAGGGAGACAAA
BACKWARD: CACTATAGGCGCAATTGGGTA
Insert Length: 963 Std Error: 0.00
Plate: CL000345 row: H column: 08
Seq primer: CCCTCACTAAGGGAGACAAA
POLYA=Yes.
```

```
Location/Qualifiers
1..963
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBUW:25134"
/sex="male"
/tissue_type="blood"
/cell_type="PBMC"
```

```
FEATURES
source
```

```
/dev_stage="adult"
/lab_host="Electromax DH108"
/clone_lib="Katze_MRPB2"
/notes="Vector: pDONR 222; Site 1: BsrG I; Site 2: BsrG I;
Created from CloneMiner cDNA Library Construction kit
(catalog #18249-029)"

ORIGIN

Query Match 93.0%; Score 831.8; DB 7; Length 963;
Best Local Similarity 97.8%; Pred. No. 3.3e-201;
Matches 864; Conservative 0; Mismatches 17; Indels 2; Gaps 2;

Qy 14 GGGGTACACAGATGCGACCGGACCGGTGGAGCTCCATAGCTTAAGCTTGCAGCACTAA 73
Db 3 GGGGTACACAGATGCGACCGGACCGGTGGAGCTCCATAGCTTGAAGCTTGCAGCACTAA 62
Qy 74 AGCAAGAATGTCCTGCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACA 133
Db 63 AGCAAGAATGTCCTGCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACA 122
Qy 134 GACTCCAGGCATATCTTGAAGAAATGCTGGAAGAGGAGGCAAAATGAAGAGATGTACTGG 193
Db 123 GACTCCAGGCATATCTTGAAGAAATGCTGGAAGAGGAGGCAAAATGAAGAGATGTACTGG 182
Qy 194 GAGATGAAACAGAGGAGAGAGAAACCAAGCCCATTTGAGCTCCCTGTCACAAAGAGCAAC 253
Db 183 GAGATGAAACAGAGGAGAGAGAAACCAAGCCCATTTGAGCTTACCTGTCCAAAGAGAGCAAC 242
Qy 254 CCCTTGAAACCACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAAAATTTACATCTGAAA 313
Db 243 CCCTTGAAACCACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAAAATTTACATCTGAAA 302
Qy 314 TACCACAGACTGAGAGAAATGCAAGAGAGGCTGAAAGATTTCAATGTACCTGTGAGCTTGG 373
Db 303 TACCACAGACTGAGAGAAATGCAAGAGAGGCTGAAAGATTTCAATGTACCTGTGAGCTTGG 362
Qy 374 AGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAGGTC 433
Db 363 AGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAGGTC 422
Qy 434 TGTCATCTGATTAACAAACCTTATGTTTAACTTGGATAAGCTGGAAGGAAAGAGCTCAAGAT 493
Db 423 TGTCATCTGATTAACAAACCTTATGTTTAACTTGGATAAGCTGGAAGGAAAGAGCTCAAGAT 482
Qy 494 TTGGTTTGAATGTCTCTCAATCTCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGA 553
Db 483 TTGGTTTGAATGTCTCTCAATCTCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGA 542
Qy 554 GGAAGAGCGATTTGGGATTTGTCAAGTTTCAAGTTCACTGGAACCTGGAACCCACAGAGATACAG 613
Db 543 GGAAGAGCGATTTGGGATTTGTCAAGTTTCAAGTTCACTGGAACCTGGAACCCACAGAGATACAG 602
Qy 614 AGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 673
Db 603 AGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
Qy 674 CTTTCTGTTCTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 733
Db 663 CTTATCTGTTCTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 722
Qy 734 TGACAGTCAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 793
Db 723 TGACAGTCAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 782
Qy 794 CATGAACCTGGCAGCAGTTTGAATTTATTTGCTGTTTCAAGCTTTAAGTTTGTGTTTGTG 852
Db 783 CGTGAACCTGGCAGCAGTTTGAATTTATTTGCTGTTTCAAGCTTTAAGTTTGTGTTTGTG 842
Qy 853 TGTTTGTGTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 894
Db 843 TGTTTGTGTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 885
```



```
RESULT 8
BM559381
LOCUS
DEFINITION AGENCOURT_659403 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5474488
5', mRNA sequence.
ACCESSION BM559381
VERSION BM559381.1 GI:18802920
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1030)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTB/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1987 row: m column: 17
High quality sequence stop: 620.
FEATURES
Location/Qualifiers
1..1030
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5474488"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="PH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
ORIGIN
Query Match 92.6%; Score 828; DB 4; Length 1030;
Best Local Similarity 97.6%; Pred. No. 3.2e-200;
Matches 851; Conservative 0; Mismatches 20; Indels 1; Gaps 1;
QY 22 AAGATGGCGACGACGCGTGGAGCTCCATAAGCTTAAGCTTGGCCGAACCTAAAGCAAGAA 81
DB 1 AAGATGGCGACGACGCGTGGAGCTCCATAGCTTAAGCTTGGCCGAACCTAAAGCAAGAA 60
QY 82 TGTCTTGTCTCGTGGTTTGGAGACCAAGGAATAAAGCAAGATCTTATCCACAGACTCCAG 141
DB 61 TGTCTTGTCTCGTGGTTTGGAGACCAAGGAATAAAGCAAGATCTTATCCACAGACTCCAG 120
QY 142 GCATATCTTGAAGAACATGCTGAAGAGGAGGCGCAATGAAGAAGATGCTATGGGAGATGAA 201
DB 121 GCATATCTTGAAGAACATGCTGAAGAGGAGGCGCAATGAAGAAGATGCTATGGGAGATGAA 180
QY 202 ACAGAGGAAGAGAAACAAAGCCATTGAGCTCCCTGTCAAGAGAGGAAGAACCCCTGAA 261
DB 181 ACAGAGGAAGAGAAACAAAGCCATTGAGCTCCCTGTCAAGAGAGGAAGAACCCCTGAA 240
QY 262 AAAAAGCTTTGATGTGGCAGCAGAGAAAGAGTGGTGAAATTTACATCTGAAATACACAG 321
DB 241 AAAAAGCTTTGATGTGGCAGCAGAGAGAAAGTGGTGAAATTTACATCTGAAATACACAG 300
QY 322 ACTGAGAGAAATGAGAGAGGGCTGAACGATTCATGTAACCTGCTGAGCTTGGAGAGTAAG 381
DB 301 ACTGAGAGAAATGAGAGAGGGCTGAACGATTCATGTAACCTGCTGAGCTTGGAGAGTAAG 360
```

```
QY 382 AAGAGCTGCTCGGCGACGCTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAAGGCTGTCTCATCT 441
DB 361 AAGAGCTGCTCGGCGACGCTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAAGGCTGTCTCATCT 420
QY 442 GATAACAACCTATGTTTAACTTGCATGAAGCTGAAGGAAAGAGCTCAAAAGATTGCTGTTTG 501
DB 421 GATAACAACCTATGTTTAACTTGCATGAAGCTGAAGGAAAGAGCTCAAAAGATTGCTGTTTG 480
QY 502 AATGTCCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAAGGAG 561
DB 481 AATGTCCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAAGGAG 540
QY 562 CGATTTGGATTTGTCACAAAGTTTCACTGGAAGTGAACCAACAGAGGATACAGAGGCAAG 621
DB 541 CGATTTGGATTTGTCACAAAGTTTCACTGGAAGTGAACCAACAGAGGATACAGAGGCAAG 600
QY 622 AAGAGGAAAGAGCAGAGCGCTTTGGGATTCGCTGATGAAAGTTCCTGATACATCTTCTGT 681
DB 601 AAGAGGAAAGAGCAGAGCGCTTTGGGATTCGCTGATGAAAGTTCCTGATACATCTTCTGT 660
QY 682 TCTCAGTGTGTTTCCATTTCTCTCCTTCTTCTTGGTCAATATATGCTTAAATGCACAGT 741
DB 661 TCTCAGTGTGTTTCCATTTCTCTCCTTCTTCTTGGGACATATATGCTTAAATGCACAGT 720
QY 742 CATGTCCTTACGCTCTGCTCGCAATGAGGAGGAGCATGTACCCAGGATACATCCATGAAGT 801
DB 721 CAGGTGCTCTACCGTCTGCTCGCAATGAGGAGGAGCATGTACCCAGGATACATCCATGAAGT 780
QY 802 GCGGAGCAGATTTGACATTTATTTGCTGTTTCAAGCTTTAAGGTTGCTGTTGTTTGTGA 861
DB 781 GCGGAGCAGATTTGACATTTATTTGCTGTTTCAAGCTTTAAGGTTGCTGTTGTTTGTGA 840
QY 862 TT-ATGTTGCTTGTGTTTAAATAAAAAAATAGAA 892
DB 841 TTAAGTGTGCTTGTGTTTAAATAAAAAAATAGAA 872
RESULT 9
LOCUS
DEFINITION AGENCOURT_7994057 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6082078
5', mRNA sequence.
ACCESSION BUI74287
VERSION BUI74287.1 GI:22688271
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2310 row: a column: 23
High quality sequence stop: 650.
FEATURES
Location/Qualifiers
1..895
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6082078"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
```


364 GAGTAAAGAGTGTCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTTCAACCAAAAGGTTCT 423
435 GTCACTGTGATAACAAACCTTATGTTAACTTGTGATAGCTTGAAGGAAAGAGCTCAAAAGATT 494
424 GTCACTGTGATAACAAACCTTATGTTAACTTGTGATAGCTTGAAGGAAAGAGCTCAAAAGATT 483
495 TGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTCAAGATGATGAGAACTGAAAAAGAG 554
484 TGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTCAAGATGATGAGAACTGAAAAAGAG 543
555 GAAGAGCGATTTGGGATTTGTACAAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTT 614
544 GAAGAGCGATTTGGGATTTGTACAAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTT 603
615 GCACAAAGAGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAGATTCTCTGATAC 674
604 GCACAAAGAGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAGATTCTCTGATAC 663
675 TTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 734
664 TTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 723
735 GCACAGTATGTGCTTACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 794
724 GCACAGTATGTGCTTACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 783
795 ATGAAGTGGGAGAGAG-TTTCACTTATTGCT-ATTTCAGCTTTTAAGG--TTGTTGTGTT 850
784 ATGAAGTGGGAGAGAGTTTGAACCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 843
851 TTTGTTTGTGATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 887
844 TTTGTTTGTGATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 880

RESULT 11
BU855435
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BU855435 928 bp mRNA linear EST 16-OCT-2002
AGENCOURT 10457152 NIH_MGC_109 Homo sapiens cdna clone
IMAGE:664932 5', mRNA sequence.
BU855435
BU855435.1 GI:24040401
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cdna library Preparation: Rubin Laboratory
cdna library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2881 row: f column: 04
High quality sequence start: 32
High quality sequence stop: 702.
Location/Qualifiers
1. .928
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:664932"
/issue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"

/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN

Query Match 91.7%; Score 820.2; DB 5; Length 928;
Best Local Similarity 96.7%; Pred. No. 3e-198;
Matches 848; Conservative 0; Mismatches 25; Indels 4; Gaps 1;

QY 21 CAAGATGCGGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGCAGCAATAAGCAAGA 80
DB 35 CACGAGGCGGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGCAGCAATAAGCAAGA 94
QY 81 ATGCTCTGCTCGTGGTGGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCA 140
DB 95 ATGCTCTGCTCGTGGTGGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCA 154
QY 141 GGCATATCTTGAAGAACATGCTGAAGAGAGGCGCAATGAAGAGATGTACTGGGAGATGA 200
DB 155 GGCATATCTTGAAGAACATGCTGAAGAGAGGCGCAATGAAGAGATGTACTGGGAGATGA 214
QY 201 AACGAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTGGA 260
DB 215 AACGAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTGGA 274
QY 261 AAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAAGATTTACATCTGAATACCA 320
DB 275 AAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAAGATTTACATCTGAATACCA 334
QY 321 GACTGAGAGAATGCAGAGAGGCGCTGAACGATTTCAATGCTGCTGAGCTGGAGAGTAA 380
DB 335 GACTGAGAGAATGCAGAGAGGCGCTGAACGATTTCAATGCTGCTGAGCTGGAGAGTAA 394
QY 381 GAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCACTTCCACAAAGAGTCTGTCTATC 440
DB 395 GAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCACTTCCACAAAGAGTCTGTCTATC 454
QY 441 TGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGAGCTCAAGATTTGGTTT 500
DB 455 TGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGAGCTCAAGATTTGGTTT 514
QY 501 GAAAGCTCTCTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAGAGAGGAAGA 560
DB 515 GAAAGCTCTCTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAGAGAGGAAGA 574
QY 561 GCGATTTGGGATTTGTCAACAGTTTCAAGCTGGAACCTGGAACCAAGAGGATACAGAGGCA 620
DB 575 GCGATTTGGGATTTGTCAACAGTTTCAAGCTGGAACCTGGAACCAAGAGGATACAGAGGCA 634
QY 621 GAAGAGGAAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTCTCTGATACCTTCTG 680
DB 635 GAAGAGGAAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTCTCTGATACCTTCTG 694
QY 681 TTCTCCAGTGTCTTCCATTTCT 740
DB 695 TTCTCCAGTGTCTTCCATTTCT 754
QY 741 TCATGCTGCTACGCTCCTGCTCGCAATGAGGAGGATGTACCCAGGATACATCCATGAAC 800
DB 755 TCATGCTGCTACGCTCCTGCTCGCAATGAGGAGGATGTACCCAGGATACATCCATGAAC 814
QY 801 TCGCGCAGCAGTTTGACTTATTTGCTGTTTCAAGTTTAA-----GGTGTGTGTTTGTGTT 856
DB 815 TCGCGCAGCAGTTTGACTTATTTGCTGTTTCAAGTTTAAAGGTTGGTGGGTTTGTGTTT 874
QY 857 TTTGATTATGTTGCTTGTGTTTAAATAAATAAGAAA 893
DB 875 TTTGATTATGTTGCTTGTGTTTAAATAAATAAGAAA 911

```

RESULT 12
LOCUS      BX456776
DEFINITION BX456776 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004YA11 linear mRNA EST 06-MAY-2004
5-PRIME, mRNA sequence.
ACCESSION  BX456776
VERSION     BX456776.2 GI:47072683
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 843)
            Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
            Full-length cDNA libraries and normalization
            Unpublished (2001)
            On May 22, 2003 this sequence version replaced gi:31036705.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 6407.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0CAP004AA06QPL&c=6407.r.

FEATURES             source
            1..843
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0CAP004YA11"
            /tissue_type="THYMUS"
            /clone_lib="Homo sapiens THYMUS"
            /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
            with a NotI-oligo(dT) primer. Five prime end enriched,
            double-strand cDNA was digested with Not I and cloned into
            the Not I and EcoRV sites of the pCMVSPORT 6 vector.
            Library was not normalized."

ORIGIN
Query Match      91.5%; Score 817.6; DB 5; Length 843;
Best Local Similarity 99.4%; Pred. No. 1.4e-197;
Matches 831; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 59 AGCTTGCAGAACTAAAGCAAGAAATGTCTTGTCTGCTGGTTTGAGACCCAGGGAATAAAGC 118
DB 1 AGCTTGCAGAACTAAAGCAAGAAATGTCTTGTCTGCTGGTTTGAGACCCAGGGAATAAAGC 60
QY 119 AGATCTTATCCACAGACTCCAGGCAATATCTTGAAGACATCTCTGAAGAGGAGGCAATG 178
DB 61 AGATCTTATCCACAGACTCCAGGCAATATCTTGAAGACATCTCTGAAGAGGAGGCAATG 120
QY 179 AAGAAGATGATCTGGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTCAGCTCCCTG 238
DB 121 AAGAAGATGATCTGGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTCAGCTCCCTG 180
QY 239 TCAAGAGAGAAAGACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAAAGTGGTGA 298
DB 181 TCAAGAGAGAAAGACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAAAGTGGTGA 240
QY 299 AAATTACATCTCAATATACACAGACTGAGAGATGCAGAGAGGCTGAACGATTCAATG 358
DB 241 AAATTACATCTCAATATACACAGACTGAGAGATGCAGAGAGGCTGAACGATTCAATG 300
QY 359 TACCTGTGAGCTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAG 418
DB 301 TACCTGTGAGCTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAG 360

```

```

QY 419 TTCCAACAAAAGGCTCTGTCTCATTAACAAAACCTATGGTTAACTTGGATTAAGCTGAAGG 478
DB 361 TTCCAACAAAAGGCTCTGTCTCATTAACAAAACCTATGGTTAACTTGGATTAAGCTGAAGG 420
QY 479 AAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTCAATCTCCAGAAAGTCTCGAAGATGATG 538
DB 421 AAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTCAATCTCCAGAAAGTCTCGAAGATGATG 480
QY 539 AGAAACTGAAAAGAGAGAGGAGCGATTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAA 598
DB 481 AGAAACTGAAAAGAGAGAGGAGCGATTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAA 540
QY 599 CCACAGAGGATACAGAGGCAAGAGAGAAAGAGAGAGAGAGAGAGCGCTTTGGGATTCGCTGAT 658
DB 541 CCACAGAGGATACAGAGGCAAGAGAGAAAGAGAGAGAGAGAGAGCGCTTTGGGATTCGCTGAT 600
QY 659 GAAAAGTTCCTGATACACTTCTGTTCTCCAGTGTTTCCATTCTCTCTCTCTCTCTCTCTCT 718
DB 601 GAAAAGTTCCTGATACACTTCTGTTCTCCAGTGTTTCCATTCTCTCTCTCTCTCTCTCTCT 660
QY 719 ACATATATGCTTAATGCACAGTCATGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 778
DB 661 ACATATATGCTTAATGCACAGTCATGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 720
QY 779 TACCCAGGATACATCCATGAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 838
DB 721 TCCCCAGGATACATCCATGAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
QY 839 GGTGTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 894
DB 780 GGTGTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 835

RESULT 13
LOCUS      BX456775/c
DEFINITION BX456775 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004YA11 linear mRNA EST 06-MAY-2004
3-PRIME, mRNA sequence.
ACCESSION  BX456775
VERSION     BX456775.2 GI:47071640
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 817)
            Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
            Full-length cDNA libraries and normalization
            Unpublished (2001)
            On May 22, 2003 this sequence version replaced gi:31034801.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 6407.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0CAP004AA06NPL&c=6407.r.

FEATURES             source
            1..817
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0CAP004YA11"
            /tissue_type="THYMUS"
            /clone_lib="Homo sapiens THYMUS"
            /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
            with a NotI-oligo(dT) primer. Five prime end enriched,
            double-strand cDNA was digested with Not I and cloned into
            the Not I and EcoRV sites of the pCMVSPORT 6 vector.
            Library was not normalized."

```

double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN		Query Match									
		Best Local Similarity									
		Matches 810; Conservative 5; Mismatches 2; Indels 0; Gaps 0;									
QY	59	AGCTTCCCGAATCTAAAGCAAGATGCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAAGC	118								
Db	817	AGCTTCCCGAATCTAAAGCAAGATGCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAAGC	758								
QY	119	AGATCTTATCCACAGACTCCAGGCATCTTTGAAGAACATGCTGAAGAGGAGGCAATG	178								
Db	757	AGATCTTATCCACACACTCCAGGCATCTTTGAAGAACATGCTGAAGAGGAGGCAATG	698								
QY	179	AGAAGATGCTACTGGGAGATGAACAGAGAGAGAGAGAACCAAGCCCATTTGACTCCCTG	238								
Db	697	AGAAGATGCTACTGGGAGATGAACAGAGAGAGAGAGAACCAAGCCCATTTGACTCCCTG	638								
QY	239	TCAAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGA	298								
Db	637	TCAAAGAGAGAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGA	578								
QY	299	AAATTACATCTGAATPACACAGACTGAGAGATGAGAGAGGCTGAACGATTCATG	358								
Db	577	AAATTACATCTGAATPACACAGACTGAGAGATGAGAGAGGCTGAACGATTCATG	518								
QY	359	TACCTGTGAGCTTGGAGATGAAGAGCTGCTGGGCAGCTAGTTTGGATTTCTTCAG	418								
Db	517	TACCTGTGAGCTTGGAGATGAAGAGCTGCTGGGCAGCTAGTTTGGATTTCTTCAG	458								
QY	419	TTCCAAACAAAGCTCTGTCATCTGATAACAAACCTATGTTTAACTTGGATAAGCTGAAGG	478								
Db	457	TTCCAAACAAAGCTCTGTCATCTGATAACAAACCTATGTTTAACTTGGATAAGCTGAAGG	398								
QY	479	AAAGAGCTCAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATG	538								
Db	397	AAAGAGCTCAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATG	338								
QY	539	AGAACTGAAAAGAGGAAGGAGCGATTTGGATTTGTACAGATTCAGCTGGAACCTGAA	598								
Db	337	AGAACTGAAAAGAGGAAGGAGCGATTTGGATTTGTCAAGATTCAGCTGGAACCTGAA	278								
QY	599	CCACAGAGATACAGAGGCAAGAGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	658								
Db	277	CCACAGAGATACAGAGGCAAGAGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	218								
QY	659	GAAGAATTCCTGATCTTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCT	718								
Db	217	GAAGAATTCCTGATCTTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCT	158								
QY	719	ACATATATGCTTAATGACAGTCTGCTACGCTACGCTGCTGCTGCTGCTGCTGCTGCT	778								
Db	157	ACATATATGCTTAATGACAGTCTGCTACGCTACGCTGCTGCTGCTGCTGCTGCTGCT	98								
QY	779	TACCCAGGTACATCATCACTGCGGAGAGAGTGTGACTTATGCTGTTTTCAGCTTTAA	838								
Db	97	TACCCAGGTACATCATCACTGCGGAGAGAGTGTGACTTATGCTGTTTTCAGCTTTAA	38								
QY	839	GGTTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT	875								
Db	37	GGTTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT	1								

RESULT 14
BU157949
LOCUS
DEFINITION
AGENCY 8118912 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6253123
5', mRNA sequence.
ACCESSION
BU157949
VERSION
BU157949.1 GI:22671859

KEYWORDS		EST.									
SOURCE		Homo sapiens (human)									
ORGANISM		Homo sapiens									
		Mammalia; Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
REFERENCE		1 (bases 1 to 910)									
AUTHORS		NIH-MGC http://mgs.nci.nih.gov/.									
TITLE		National Institutes of Health, Mammalian Gene Collection (MGC)									
JOURNAL		Unpublished (1999)									
COMMENT		Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: DCTD/DTF cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCN2400 row: h column: 20 High quality sequence stop: 645.									
FEATURES		Location/Qualifiers									
		1..910									
		/organism="Homo sapiens"									
		/mol_type="mRNA"									
		/db_xref="taxon:9606"									
		/clone="IMAGE:6253123"									
		/issue type="melanotic melanoma, cell line"									
		/lab host="DH10B (phage-resistant)"									
		/clone lib="NIH MGC 112"									
		/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."									
ORIGIN		Query Match									
		Best Local Similarity									
		Matches 822; Conservative 0; Mismatches 16; Indels 1; Gaps 1;									
QY	51	TAAGCTAAAGCTTCCCGAATCTTCTGCTCGTGGTTGGAGACCAAGG	110								
Db	1	TAAGCTAAAGCTTCCGAACTTAAAGCAAGATGCTTCTGCTGGTGGAGACCAAGG	60								
QY	111	AATAAGCAAGATCTTATCCAGACTCCAGCATATCTTGAAGAACATGCTGAAGAGA	170								
Db	61	AATAAGCAAGATCTTATCCAGACTCCAGCATATCTTGAAGAACATGCTGAAGAGA	120								
QY	171	GGCAATCAAGAGATGCTGGGAGATGAAACAGAGAGAGAGAGAGAGAGAGAGAGAG	230								
Db	121	GGCAATCAAGAGATGCTGGGAGATGAAACAGAGAGAGAGAGAGAGAGAGAGAGAG	180								
QY	231	GCTCCCTCTCAAAG	290								
Db	181	GCTCCCTCTCAAAG	240								
QY	291	AGTGGTGAATTTACATCTGAAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	350								
Db	241	AGTGGTGAATTTACATCTGAAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	300								
QY	351	ATTCATCTGCTGAGCTTGGAGAGTAAAGAGCTCTCGGCAGCTAGCTTTGGGAT	410								
Db	301	ATTCATCTGCTGAGCTTGGAGAGTAAAGAGCTCTCGGCAGCTAGCTTTGGGAT	360								
QY	411	TTCTTCAGTCCCAACAAAGGCTCTGCTCATCTGATAACAAACCTATGTTAACTTGAATA	470								
Db	361	TTCTTCAGTCCCAACAAAGGCTCTGCTCATCTGATAACAAACCTATGTTAACTTGAATA	420								
QY	471	GCTGAAGAGAGAGCTCAAAGATTTGGTTTGAATGCTCTCTCAATCTCCAGAAAGTCTGA	530								

421	Db	GCTGAAAGGAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTCTCAATCTCAGAAAGTCTGA	480
531	Qy	AGATGATGAGAAAATCGAAAAAGAGGAGGACCGATTTGGGATTGTCACAAGTTCAGCTGG	590
481	Db	AGATGATGAGAAAATCGAAAAAGAGGAGGACCGATTTGGGATTGTCACAAGTTCAGCTGG	540
591	Qy	AAC TTGGAACCA CAGAGGATACAGAGGCCAAAGAGAGGAAAAAGACAGAGCGCTTTGGGAT	650
541	Db	AAC TTGGAACCA CAGAGGATA CAGAGGCCAAAGAGAGGAAAAAGACAGAGCGCTTTGGGAT	600
651	Qy	TGCCTGATGAAAAGTTCTTGATACTTCTCTGTCTTCCAGTGTGTTTCCATTCTCTCTCTCT	710
601	Db	TGCCTGATGAAAAGTTCTTGATACTTCTCTGTCTTCCAGTGTGTTTCCATTCTCTCTCTCT	660
711	Qy	TC TTGGTGCACATATATGCTCTAAATGCA CAGTCATGTGCTCAGTCCTGCCTCGCAATGAG	770
661	Db	TC TTGGTGCACATATATGCTCTAAATGCA CAGTCATGTGCTCAGTCCTGCCTCGCAATGAG	720
771	Qy	GGAGCATGTATCCCCAGGTACATCCATGAATCGCGGCAGCATTTTGACTTATTTGCTGTTTC	830
721	Db	GGAGCATGTATCCCCAGGTACATCCATGAATCGCGGCAGCATTTTGACTTATTTGCTGTTTC	780
831	Qy	AGCTTTTAAGGTG-TTGTTGTTTTGTTTTTTCGATTATGTTGCTGTATATAAAAAAAT	888
781	Db	AGCTTTTAAGGTGTTTGGGGTTTTGTTTTGATAGGTGGCTCGGTACCAAAACAAGATT	839

RESULT 15	
BU599301	
LOCUS	823 bp mRNA linear EST 20-SEP-2002
DEFINITION	ACGENCOURT 8908959 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6458323
	5', mRNA sequence.
ACCESSION	BU599301
VERSION	BU599301.1 GI:23251060
KEYWORDS	EST.
SOURCE	Homo sapiens
ORGANISM	Homo sapiens (human)
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 823)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs- kemail.nih.gov Tissue Procurement: NCI cDNA Library Preparation: Michael Brownstein Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

5'-ATTCTAGAGCCGAGCGCGACATG-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH_MGC_141). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH MGC Library."

ORIGIN

Query Match	89.6%	Score	800.8	DB	5	Length	823	
Best Local Similarity	99.5%	Pred. No.	2.6e-193					
Matches	813	Conservative	0	Mismatches	3	Indels	1	
Qy	4	AGTGGAGT	GAGGGTAA	CAAGATG	GCGCAC	CGACAGCGTGGAGCTT	CCATAAGCTAAAGCTT	63
Db	6	AGTGGAGT	GAGGGTAA	CAAGATG	GCGCAC	CGACAGCGTGGAGCTT	CCATAAGCTTAAAGCTT	65
Qy	64	GCCGAAC	TAAAGCA	GAATGTTCTT	GTGCTCGTGGTTT	GGAGACCAAGGGAAT	TAAGCAAGAT	123
Db	66	GCCGAAC	TAAAGCA	GAATGTTCTT	GTGCTCGTGGTTT	GGAGACCAAGGGAAT	TAAGCAAGAT	125
Qy	124	CTTATCC	ACAGACT	CCAGGCATATCTT	TGAAGACATGCT	TGAAGAGAGGCAAAAT	GAAGAA	183
Db	126	CTTATCC	ACAGACT	CCAGGCATATCTT	TGAAGACATGCT	TGAAGAGAGGCAAAAT	GAAGAA	185
Qy	184	GATGTAT	CTGGGAGAT	GAAACAG	AGAGGAAGAAACAAAG	CCCCATTTGAGCTCCCTGT	CAAA	243
Db	186	GATGTAT	CTGGGAGAT	GAAACAG	AGAGGAAGAAACAAAG	CCCCATTTGAGCTCCCTGT	CAAA	245
Qy	244	GAGGAGA	ACCCCTG	AAAAAACTGTT	TGATGCTGGCAGCAGAGAA	GAAGTGGTGAANA	TT	303
Db	246	GAGGAGA	ACCCCTG	AAAAAACTGTT	TGATGCTGGCAGCAGAGAA	GAAGTGGTGAANA	TT	305
Qy	304	ACATCTG	AAATATCA	CACAGACT	GAGAGAATGC	AGAGAGGGCTGAA	CGATTTCAATGTACCT	363
Db	306	ACATCTG	AAATATCA	CACAGACT	GAGAGAATGC	AGAGAGGGCTGAA	CGATTTCAATGTACCT	365
Qy	364	GTGAGCT	TGGAGAGT	TAAGAAAGCTGCT	CGGGCAGCTAG	GGTTTGGGATTTCTT	CAGTTCCA	423
Db	366	GTGAGCT	TGGAGAGT	TAAGAAAGCTGCT	CGGGCAGCTAG	GGTTTGGGATTTCTT	CAGTTCCA	425
Qy	424	ACAAAAG	CTGTG	CTATGATA	CAAAACCTTAT	GGTTTAACTTTGGAT	TAAGCTGAAGGAAAGA	483
Db	426	ACAAAAG	CTGTG	CTATGATA	CAAAACCTTAT	GGTTTAACTTTGGAT	TAAGCTGAAGGAAAGA	485
Qy	484	GCTCAAA	AGATTTG	GGTTGAAT	GTCTCTTCAAT	CTCCAGAAAGTCTG	AAAGATGATGAGAA	543
Db	486	GCTCAAA	AGATTTG	GGTTGAAT	GTCTCTTCAAT	CTCCAGAAAGTCTG	AAAGATGATGAGAA	545
Qy	544	CTGAAA	AAGAGG	AGAGCGATTT	TGGGATTTGTC	ACAAAGTTTCA	GCTGGACTGCAACCACA	603
Db	546	CTGAAA	AAGAGG	AGAGCGATTT	TGGGATTTGTC	ACAAAGTTTCA	GCTGGACTGCAACCACA	605
Qy	604	GAGGATA	CAGAGCA	AGAGGAAGAGG	AAAAAGAGCAG	AGCGCTTTGGGATTTG	CGCTCATGAAAA	663
Db	606	GAGGATA	CAGAGCA	AGAGGAAGAGG	AAAAAGAGCAG	AGCGCTTTGGGATTTG	CGCTCATGAAAA	665
Qy	664	GTTCCTG	ATACTTTCT	GTTCTCCAGT	TTTTTCCATTTCT	CTCTCTTCTTCTG	GTGCACATA	723
Db	666	GTTCCTG	ATACTTTCT	GTTCTCCAGT	TTTTTCCATTTCT	CTCTCTTCTTCTG	GTGCACATA	725
Qy	724	TATGCCT	TAAATG	CACAGT	CTATGTG	CCCTACGCTCG	CAATGAGGAGCATGTACCC	783
Db	726	TATGCCT	TAAATG	CACAGT	CTATGTG	CCCTACGCTCG	CAATGAGGAGCATGTACCC	785
Qy	784	CAGGATAC	TCCAT-G	AATCTGGG	CAGCAGTTT	GACTTT	819	
Db	786	CAGGATAC	TCCATG	GAATCT	GGCAGCAGTTT	GACTTT	822	

Search completed: October 5, 2005, 00:04:19
Job time : 3579.42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 00:04:33 ; Search time 97 Seconds
(without alignments)
837.316 Million cell updates/sec

Title: US-09-788-476A-2

Perfect score: 1040

Sequence: 1 MATETVELHLKLAELKQEC.....GTTEDTEAKRKRAERFGIA 210

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	1040	100.0	210	4 AAB36609	Human FLE
2	1040	100.0	210	5 ABB97210	Novel hum
3	1040	100.0	210	5 ABG72881	Novel hum
4	1040	100.0	228	4 AAU18233	Novel hum
5	1040	100.0	228	5 ABG92654	Human DNA
6	1040	100.0	228	7 ADC25371	Human ext
7	746	71.7	185	8 ADF30022	Human sec
8	742	71.3	149	3 AAG00732	Human sec
9	222	21.3	308	4 ABB63720	Drosophil
10	155	14.9	734	7 ADE14358	Human int
11	155	14.9	747	8 ADS10484	Human the
12	155	14.9	795	4 ABB61567	Drosophil
13	139	13.4	185	3 AAG20995	Arabidops
14	139	13.4	214	3 AAG20700	Arabidops
15	138	13.3	750	8 ADR08732	Human pro
16	138	13.3	1150	4 AAM40294	Human pol
17	137.5	13.2	540	4 AAG74992	Human col
18	137.5	13.2	1150	5 ABG32465	Human pro
19	137.5	13.2	1150	8 ADS88261	Human pro
20	136.5	13.1	718	4 AAB92973	Human pro
21	136.5	13.1	718	8 ADS88419	Human pro
22	136.5	13.1	1150	7 ADF50152	Human dea
23	133.5	12.8	633	3 AAG31342	Arabidops
24	133.5	12.8	633	8 ADN72455	Thale cre
25	133	12.8	1146	7 ADF50150	Murine de

ALIGNMENTS

RESULT 1

AAB36609

ID AAB36609 standard; protein; 210 AA.

XX AAB36609;

XX 09-MAR-2001 (first entry)

DE Human FLEXHT-31 protein sequence SEQ ID NO:31.

XX Human; FLEXHT: full-length molecules expressed in human tissue;
diagnosis; gene expression; genetic linkage; genetic variability;
cytotoxic; anticonvulsant; antiarteriosclerotic; immunomodulatory;
antitumor; antitumor; antitumor; antitumor; antitumor; antitumor;
anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;
antipsoriatic; antirheumatic; antiulcer; gene therapy; anaemia; gout;
epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;
cancer; immunological disorder; asthma; bronchitis; cirrhosis;
Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;
osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;
ulcerative colitis.

XX Homo sapiens.

XX WO200070047-A2.

XX 23-NOV-2000.

XX 12-MAY-2000; 2000WO-US013299.

XX 14-MAY-1999; 99US-00311894.

XX 14-MAY-1999; 99US-00311937.

XX 14-MAY-1999; 99US-00311940.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Tang YT, Lal P, Reddy R, Batra S, Baughn MR, Yang J;
Azimzai Y, Lu DAM, Au-Young J, Shih LL;

XX WPI; 2001-016234/02.

XX N-PSDB; AAC88100.

XX Human FLEXHT protein and DNA sequences, useful for treating immunological
disorders, developmental disorders, and cancers.

XX Claim 1; Page 120; 168pp; English.

XX AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules

CC

CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The
 CC present invention describes an isolated polypeptide (A) comprising an
 CC amino acid sequence selected from one of 55 amino acid sequences 42-876
 CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %
 CC identical sequence, and a biologically active or immunogenic fragment of
 CC the sequence. The FLEXHT proteins can have antianaemic, anticonvulsant,
 CC antiarteriosclerotic, immunomodulatory, cytostatic, antiaesthetic,
 CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antithyroid,
 CC neuroprotective, antiarthritic, osteopathic, antipsoriatic, antitumor and
 CC antirheumatic activities, and can be used in gene therapy. The
 CC polynucleotide sequences can be used to express the protein sequences.
 CC Pharmaceutical compositions comprising FLEXHT can be used to treat
 CC diseases or conditions associated with altered expression of functional
 CC FLEXHT. The proteins and polynucleotides can be used to diagnose and
 CC treat disorders including anaemia, epilepsy, arteriosclerosis,
 CC atherosclerosis, developmental disorders, cancers, and immunological
 CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,
 CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,
 CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and
 CC ulcerative colitis
 XX
 SQ Sequence 210 AA;

Query Match 100.0%; Score 1040; DB 4; Length 210;
 Best Local Similarity 100.0%; Pred. No. 1.5e-85;
 Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATETVELHLKLAELKQECCLARGLETGKIQDLIHLRLQAYLEHAEANEEDVLGDET 60
 DB 1 MATETVELHLKLAELKQECCLARGLETGKIQDLIHLRLQAYLEHAEANEEDVLGDET 60

QY 61 EEEETKPIELPVKEEPEPTVDVAEKKVKVITSEIPQTERMQKRAERFNVPSLESKK 120
 DB 61 EEEETKPIELPVKEEPEPTVDVAEKKVKVITSEIPQTERMQKRAERFNVPSLESKK 120

QY 121 AAAAAFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
 DB 121 AAAAAFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180

QY 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210
 DB 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210

RESULT 2
 ABB97210
 ID ABB97210 standard; protein; 210 AA.
 AC ABB97210;
 DT 28-JUN-2002 (first entry)
 DE Novel human protein SEQ ID NO: 478.
 DE Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
 KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
 KW neuroprotective; antiparkinsonian; protein therapy; EST;
 KW expressed sequence tag.
 XX Homo sapiens.
 XX WO200222660-A2.
 FN 21-MAR-2002.
 XX 10-SEP-2001; 2001WO-US026015.
 XX 11-SEP-2000; 2000US-00659671.
 PR (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI: 2002-292408/33.
 DR N-PSDB; ABN32396.
 XX
 PT An isolated polynucleotide for treating diseases associated with its
 encoded polypeptide such as cancer and multiple sclerosis.
 XX
 PS Example 2; SEQ ID NO 478; 509pp; English.
 XX
 CC The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
 CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
 CC rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a protein of the invention
 XX
 SQ Sequence 210 AA;

Query Match 100.0%; Score 1040; DB 5; Length 210;
 Best Local Similarity 100.0%; Pred. No. 1.5e-85;
 Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATETVELHLKLAELKQECCLARGLETGKIQDLIHLRLQAYLEHAEANEEDVLGDET 60
 DB 1 MATETVELHLKLAELKQECCLARGLETGKIQDLIHLRLQAYLEHAEANEEDVLGDET 60

QY 61 EEEETKPIELPVKEEPEPTVDVAEKKVKVITSEIPQTERMQKRAERFNVPSLESKK 120
 DB 61 EEEETKPIELPVKEEPEPTVDVAEKKVKVITSEIPQTERMQKRAERFNVPSLESKK 120

QY 121 AAAAAFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
 DB 121 AAAAAFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180

QY 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210
 DB 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210

RESULT 3
 ABG72881
 ID ABG72881 standard; protein; 210 AA.
 XX
 AC ABG72881;
 DT 03-MAR-2003 (first entry)
 DE Novel human protein HCC-1.
 DE Human; hcc-1; human hepatocellular carcinoma tissue; HCC-M cell line;
 KW cancer; hepatocellular carcinoma; antisense gene therapy.
 XX Homo sapiens.
 XX US2002107190-A1.
 PN 08-AUG-2002.
 XX 21-FEB-2001; 2001US-00789476.
 XX 25-FEB-2000; 2000US-0185116P.
 PR (CHUN/) CHUNG C M.
 PA (CHAN/) CHAN L.
 PA (OUKK/) OU K.
 PA (ONGS/) ONG S.
 PA (SEOW/) SEOW T K.
 PA (LIAN/) LIANG C R.
 PA (CHOO/) CHOONG M L.


```

PA (TANL/) TAN L K.
XX
PI Chung CM, Chan L, Ou K, Ong S, Seow TK, Liang CR, Choong ML;
PI Tan LK;
XX
DR WPI; 2002-697878/75.
DR N-PSDB; ABX13935.
XX
PT New nucleic acid which is differentially expressed in human
PT hepatocellular carcinoma tissue useful for diagnosing and developing
PT therapy for hepatocellular carcinoma and related conditions.
XX
XX Claim 4; Fig 2; 23pp; English.
XX
CC The invention describes an isolated nucleic acid whose expression is
CC differential or preferential in human hepatocellular carcinoma tissue or
CC tissue from a related cancer relative to other tissue in the subject(s)
CC diagnosed with the condition. The nucleic acid is used to diagnose and
CC treat hepatocellular carcinoma and related cancers, or modulate one or
CC more activities in a cell e.g. by antisense gene therapy. This is the
CC amino acid sequence of the novel human protein HCC-1 identified from the
CC HCC-M cell and proposed to be involved in nucleic acid binding and
CC transcription control
XX
SQ Sequence 210 AA;

Query Match 100.0%; Score 1040; DB 5; Length 210;
Best Local Similarity 100.0%; Pred. No. 1.5e-85;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATETVHLKLAELKOECLARGLETGKIGKQDILHRLQAYLEHAEANEEDVLGDET 60
Db 1 MATETVHLKLAELKOECLARGLETGKIGKQDILHRLQAYLEHAEANEEDVLGDET 60

Qy 61 EEETKTEIPLVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERFNPVLSLSKK 120
Db 61 EEETKTEIPLVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERFNPVLSLSKK 120

Qy 121 AARARFGISSVPTKGLSSDNKPMVNDLKLKRAQRFGLNVSSISRKSEDEKLKKRKR 180
Db 121 AARARFGISSVPTKGLSSDNKPMVNDLKLKRAQRFGLNVSSISRKSEDEKLKKRKR 180

Qy 181 FGIVTSSAGTGTEDTEAKKKRAERFGIA 210
Db 181 FGIVTSSAGTGTEDTEAKKKRAERFGIA 210

RESULT 4
AAU18233
ID AAU18233 standard; protein; 228 AA.
XX
AC AAU18233;
XX
DT 21-NOV-2001 (first entry)
XX
DE Novel human DNA-binding protein #80.
XX
KW Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein; DNA organisation;
KW gene transcription; malignant disease; autoimmune disorder;
KW rheumatic disease; genetic abnormality; infectious disease;
KW neurological disorder; gene therapy; immunomodulatory; anti-HIV;
KW anti rheumatic; anti microbial; cytostatic.
XX
OS Homo sapiens.
XX
FN WO200155162-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001305.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR

```

```

PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-020515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-021680P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225113P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231424P.
PR 08-SEP-2000; 2000US-0231433P.
PR 08-SEP-2000; 2000US-0231441P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 25-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0234998P.
PR 27-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.

```

PR	02-OCT-2000;	2000US-0237038P.	XX	Claim 11; SEQ ID NO 218; 561pp; English.	
PR	02-OCT-2000;	2000US-0237039P.	PS	The present invention relates to the isolation of novel DNA-binding proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding for these proteins. DNA-binding proteins such as histones, chromo (chromatin organisation modifier) domain proteins, and Y-box binding proteins may contribute to diseases resulting from aberrant DNA organisation and/or gene transcription. The sequences of the invention are useful in screening assays to identify antagonists and/or agonists that may enhance or block activities mediated by DNA-binding proteins. Blockers of DNA-binding proteins may be useful in treating disorders such as malignant diseases (e.g. cancer), autoimmune disorders (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's disease). The polynucleotide sequences of the invention may also be used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding proteins. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	
PR	02-OCT-2000;	2000US-0237040P.	XX		
PR	13-OCT-2000;	2000US-0239935P.	CC		
PR	13-OCT-2000;	2000US-0239937P.	CC		
PR	20-OCT-2000;	2000US-0240960P.	CC		
PR	20-OCT-2000;	2000US-0241221P.	CC		
PR	20-OCT-2000;	2000US-0241785P.	CC		
PR	20-OCT-2000;	2000US-0241786P.	CC		
PR	20-OCT-2000;	2000US-0241787P.	CC		
PR	20-OCT-2000;	2000US-0241808P.	CC		
PR	20-OCT-2000;	2000US-0241809P.	CC		
PR	20-OCT-2000;	2000US-0241826P.	CC		
PR	01-NOV-2000;	2000US-0244617P.	CC		
PR	08-NOV-2000;	2000US-0246474P.	CC		
PR	08-NOV-2000;	2000US-0246475P.	CC		
PR	08-NOV-2000;	2000US-0246476P.	CC		
PR	08-NOV-2000;	2000US-0246477P.	CC		
PR	08-NOV-2000;	2000US-0246478P.	CC		
PR	08-NOV-2000;	2000US-0246523P.	CC		
PR	08-NOV-2000;	2000US-0246524P.	CC		
PR	08-NOV-2000;	2000US-0246525P.	XX		
PR	08-NOV-2000;	2000US-0246526P.	SQ		
PR	08-NOV-2000;	2000US-0246527P.	Sequence 228 AA;		
PR	08-NOV-2000;	2000US-0246528P.	Query Match 100.0%; Score 1040; DB 4; Length 228;		
PR	08-NOV-2000;	2000US-0246532P.	Best Local Similarity 100.0%; Pred. No. 1.7e-85;		
PR	08-NOV-2000;	2000US-0246609P.	Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
PR	08-NOV-2000;	2000US-0246610P.			
PR	08-NOV-2000;	2000US-0246611P.			
PR	08-NOV-2000;	2000US-0246613P.			
PR	17-NOV-2000;	2000US-0249207P.			
PR	17-NOV-2000;	2000US-0249208P.			
PR	17-NOV-2000;	2000US-0249209P.			
PR	17-NOV-2000;	2000US-0249210P.			
PR	17-NOV-2000;	2000US-0249211P.			
PR	17-NOV-2000;	2000US-0249212P.			
PR	17-NOV-2000;	2000US-0249213P.			
PR	17-NOV-2000;	2000US-0249214P.			
PR	17-NOV-2000;	2000US-0249215P.			
PR	17-NOV-2000;	2000US-0249216P.			
PR	17-NOV-2000;	2000US-0249217P.			
PR	17-NOV-2000;	2000US-0249218P.			
PR	17-NOV-2000;	2000US-0249244P.			
PR	17-NOV-2000;	2000US-0249245P.			
PR	17-NOV-2000;	2000US-0249264P.			
PR	17-NOV-2000;	2000US-0249265P.			
PR	17-NOV-2000;	2000US-0249297P.			
PR	17-NOV-2000;	2000US-0249299P.			
PR	17-NOV-2000;	2000US-0249300P.			
PR	01-DEC-2000;	2000US-0250160P.			
PR	01-DEC-2000;	2000US-0250391P.			
PR	03-DEC-2000;	2000US-0251030P.			
PR	05-DEC-2000;	2000US-0251988P.			
PR	05-DEC-2000;	2000US-0256719P.			
PR	06-DEC-2000;	2000US-0251479P.			
PR	08-DEC-2000;	2000US-0251856P.			
PR	08-DEC-2000;	2000US-0251868P.			
PR	08-DEC-2000;	2000US-0251869P.			
PR	08-DEC-2000;	2000US-0251989P.			
PR	08-DEC-2000;	2000US-0251990P.			
PR	11-DEC-2000;	2000US-0254097P.			
PR	05-JAN-2001;	2001US-0259678P.			
XX	(HUMA--)	HUMAN GENOME SCI INC.			
PA					
XX	Rosen CA, Barash SC, Ruben SM;				
XX	PI				
XX					
DR	WPI; 2001-465557/50.				
DR	N-PSDB; AAS29109.				
XX					
XX	Nucleic acid molecules encoding human secreted chromosomal binding proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers.				
PT					
PT					

```

PF 17-JAN-2001; 2001US-00764846.
XX
PF 31-JAN-2000; 2000US-0179065P.
XX
PR 04-FEB-2000; 2000US-0180628P.
PR
PR 28-JUN-2000; 2000US-0214886P.
PR
PR 07-JUL-2000; 2000US-0216647P.
PR
PR 07-JUL-2000; 2000US-0216880P.
PR
PR 11-JUL-2000; 2000US-0217487P.
PR
PR 11-JUL-2000; 2000US-0217496P.
PR
PR 14-JUL-2000; 2000US-0218290P.
PR
PR 26-JUL-2000; 2000US-0220963P.
PR
PR 26-JUL-2000; 2000US-0220964P.
PR
PR 14-AUG-2000; 2000US-0224518P.
PR
PR 14-AUG-2000; 2000US-0224519P.
PR
PR 14-AUG-2000; 2000US-0225267P.
PR
PR 14-AUG-2000; 2000US-0225268P.
PR
PR 14-AUG-2000; 2000US-0225270P.
PR
PR 14-AUG-2000; 2000US-0225447P.
PR
PR 14-AUG-2000; 2000US-0225757P.
PR
PR 14-AUG-2000; 2000US-0225758P.
PR
PR 22-AUG-2000; 2000US-0226868P.
PR
PR 30-AUG-2000; 2000US-0228924P.
PR
PR 01-SEP-2000; 2000US-0229287P.
PR
PR 01-SEP-2000; 2000US-0229343P.
PR
PR 01-SEP-2000; 2000US-0229344P.
PR
PR 01-SEP-2000; 2000US-0229345P.
PR
PR 05-SEP-2000; 2000US-0229509P.
PR
PR 05-SEP-2000; 2000US-0229513P.
PR
PR 08-SEP-2000; 2000US-0231413P.
PR
PR 21-SEP-2000; 2000US-0234223P.
PR
PR 21-SEP-2000; 2000US-0234274P.
PR
PR 25-SEP-2000; 2000US-0234997P.
PR
PR 27-SEP-2000; 2000US-0235834P.
PR
PR 29-SEP-2000; 2000US-0236327P.
PR
PR 29-SEP-2000; 2000US-0236367P.
PR
PR 29-SEP-2000; 2000US-0236368P.
PR
PR 29-SEP-2000; 2000US-0236369P.
PR
PR 29-SEP-2000; 2000US-0236370P.
PR
PR 02-OCT-2000; 2000US-0236802P.
PR
PR 02-OCT-2000; 2000US-0237037P.
PR
PR 02-OCT-2000; 2000US-0237038P.
PR
PR 02-OCT-2000; 2000US-0237039P.
PR
PR 03-OCT-2000; 2000US-0237040P.
PR
PR 13-OCT-2000; 2000US-0239935P.
PR
PR 20-OCT-2000; 2000US-0240960P.
PR
PR 20-OCT-2000; 2000US-0241785P.
PR
PR 20-OCT-2000; 2000US-0241809P.
PR
PR 01-NOV-2000; 2000US-0244617P.
PR
PR 17-NOV-2000; 2000US-0249299P.
PR
PR 08-DEC-2000; 2000US-0251856P.
PR
PR 08-DEC-2000; 2000US-0251868P.
PR
PR 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
PA
PA (RUBE/) RUBEN S M.
PA
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2002-690611/74.
DR
DR N-PSDB; ABS68249.
XX
XX
XX Novel DNA-binding protein useful for diagnosis, prognosis, prevention and
PT treatment of immune, hyperproliferative, respiratory, cardiovascular,
PT reproductive, endocrine, gastrointestinal and neurological disorders.
XX
XX Claim 11; SEQ ID NO 218; 225pp; English.
XX
XX
XX The present invention relates to a new DNA-binding protein. The invention
CC is useful in treating, preventing, diagnosing and/or prognosing
CC immunodeficiencies (e.g. B cell immunodeficiencies, severe combined
CC immunodeficiencies), autoimmune disorders (rheumatoid arthritis, multiple
CC sclerosis, diabetes mellitus), allergic reactions and conditions (e.g.

```

```

CC asthma), inflammatory conditions, graft-versus-host disease, blood-
CC related disorders (thrombosis, atherosclerosis), hyperproliferative
CC disorders (e.g. cancer), renal disorders (e.g. acute glomerulonephritis),
CC cardiovascular disorders (e.g. arrhythmic), respiratory disorders
CC (Goodpasture's syndrome), neurological disorders (e.g. Alzheimer's
CC disease, Parkinson's disease), endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders (e.g. endometriosis), infectious
CC diseases (e.g. viral, bacterial or fungal infections) and
CC gastrointestinal disorders (e.g. Crohn's disease). The invention is also
CC useful to stimulate neuronal growth and treat, prevent, and/or diagnose
CC neuronal damage which occurs in certain neuronal disorders or neuro-
CC degenerative conditions. The present amino acid sequence represents a
CC human DNA-binding protein of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC http.seqdata.uspto.gov/sequence
XX
SQ Sequence 228 AA;
Query Match 100.0%; Score 1040; DB 5; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.7e-85;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MATETVELHKLKLAELKQECLEARGLETGKIKODLIHRLQAYLEHAEENEEDVLGDET 60
Db |||||
Qy 61 EEEETKPIELPVKEEPPPEKTVVAAEKVKVKITSEIPQTERMQKRAERFNPVLSKK 120
Db |||||
Qy 79 EEEETKPIELPVKEEPPPEKTVVAAEKVKVKITSEIPQTERMQKRAERFNPVLSKK 138
Db |||||
Qy 121 ARAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKE 180
Db |||||
Qy 139 ARAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKE 198
Db |||||
Qy 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210
Db 199 FGIVTSAGTGTTEDTEAKKRAERFGIA 228
RESULT 6
ID ADC25371 standard; protein; 228 AA.
XX
AC ADC25371;
XX
XX 18-DEC-2003 (first entry)
XX
XX Human extracellular matrix protein from gene 80.
XX
XX Extracellular matrix protein; cytostatic; antibacterial; virucide;
XX neuroprotective; gynaecological; gastrointestinal-Gen; cardiac;
XX cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
XX respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
XX neurotropic; anti-allergic; cancer; bacterial infection; viral infection;
XX neural disorder; immune system disorder; blood disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX inflammatory disorder; proliferative disorder; human; gene therapy.
XX
XX Homo sapiens.
XX
XX OS
XX
XX US2003049650-A1.
XX
XX 13-MAR-2003.
XX
XX 07-MAR-2002; 2002US-00091483.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.

```

PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0237009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234937P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764846.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Ruben SM, Barash SC;
WPI; 2003-605749/57.
N-PSDB; ADC25243.
New DNA-binding proteins and gene encoding them, useful for diagnosing,
treating and/or preventing e.g. neurological, inflammatory, infectious,
cardiovascular, autoimmune, respiratory, neoplastic or digestive
diseases.
Claim 11; SEQ ID NO 218; 226pp; English.
PR

XX The invention relates to an isolated nucleic acid molecule (cDNA)
CC encoding a human extracellular matrix protein, representing one of 161
CC novel genes. Also included are recombinant vectors, host cells
CC (expressing the protein), the extracellular matrix proteins (including
CC their fragments, epitopes and homologues), an isolated antibody that
CC binds specifically to the protein, diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or absence of a mutation in the nucleic acid and diagnosing a
CC condition based on the presence or absence of the mutation), diagnosing a
CC pathological condition or susceptibility to a pathological condition
CC (comprising determining the presence or amount of expression of the
CC protein in a biological sample and diagnosing a condition based on the
CC presence or amount of expression of the protein), preventing, treating or
CC ameliorating a medical condition by administering the nucleic acid or
CC protein to a mammalian subject, identifying a binding partner to the
CC protein, the gene corresponding to the cDNA sequence, and identifying an
CC activity in a biological assay (comprising expressing the nucleic acid in
CC a cell, isolating the supernatant, detecting an activity in a biological
CC assay and identifying the protein in the supernatant having the
CC activity). The nucleic acids and proteins display the following
CC activities Cytostatic, antibacterial, Virucide, Neuroprotective,
CC Gynaecological, Gastrointestinal-Gen, Cardiant, Cardiovascular-Gen,

Query Match 100.0%; Score 1040; DB 7; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.7e-85;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATETVELHKLKLAELKQELCARGLETKIGIKQDILHRLQAYLEHAEANEEDVLGDET 60
Db 19 MATETVELHKLKLAELKQELCARGLETKIGIKQDILHRLQAYLEHAEANEEDVLGDET 78

Qy 61 EEEETKPIELPVKEEPEKTVDAAEKVKVITSEIPQTERMQKRAERFNVPSLESKK 120
Db 79 EEEETKPIELPVKEEPEKTVDAAEKVKVITSEIPQTERMQKRAERFNVPSLESKK 138

Qy 121 AARARFGISVPYTKGLSSDNKPMWNLDKLKERQRFGLNVSSISRKSEDEKLUKGRKER 180
Db 139 AARARFGISVPYTKGLSSDNKPMWNLDKLKERQRFGLNVSSISRKSEDEKLUKGRKER 198

Qy 181 FGIYTSAGTCTTETDAKKKRAERFGIA 210
Db 199 FGIYTSAGTCTTETDAKKKRAERFGIA 228

RESULT 7
ADP30022
ID ADP30022 standard; protein; 185 AA.
XX AC ADP30022;
XX DT 12-AUG-2004 (first entry)
XX DE Human secreted protein SEQ ID #789.
XX KW Cytostatic; Antinflammatory; Immunosuppressive; Antibacterial; Virucide;
XX KW cancer; inflammatory; immune; human secreted protein.
XX OS Homo sapiens.
XX FN WO2004035732-A2.
XX PD 29-APR-2004.
XX PF 28-AUG-2003; 2003WO-US026780.
XX PR 29-AUG-2002; 2002US-0406576P.
PR 29-AUG-2002; 2002US-0406579P.
PR 29-AUG-2002; 2002US-0406585P.
PR 29-AUG-2002; 2002US-0406586P.
PR 29-AUG-2002; 2002US-0406608P.
PR 29-AUG-2002; 2002US-0406611P.
PR 29-AUG-2002; 2002US-0406612P.

PR 29-AUG-2002; 2002US-0406616P.
PR 29-AUG-2002; 2002US-0406640P.
PR 29-AUG-2002; 2002US-0406642P.
PR 29-AUG-2002; 2002US-0406646P.
PR 29-AUG-2002; 2002US-0406653P.
PR 29-AUG-2002; 2002US-0406655P.
PR 29-AUG-2002; 2002US-0406666P.
PR 17-SEP-2002; 2002US-0410946P.
PR 17-SEP-2002; 2002US-0410947P.
PR 17-SEP-2002; 2002US-0410948P.
PR 17-SEP-2002; 2002US-0410949P.
PR 17-SEP-2002; 2002US-0410953P.
PR 17-SEP-2002; 2002US-0410957P.
PR 17-SEP-2002; 2002US-0410958P.
PR 17-SEP-2002; 2002US-0410959P.
PR 17-SEP-2002; 2002US-0410960P.
PR 17-SEP-2002; 2002US-0410961P.
PR 17-SEP-2002; 2002US-0410962P.
PR 17-SEP-2002; 2002US-0410963P.
PR 17-SEP-2002; 2002US-0411023P.
PR 17-SEP-2002; 2002US-0411024P.
PR 17-SEP-2002; 2002US-0411032P.
PR 17-SEP-2002; 2002US-0411035P.
PR 17-SEP-2002; 2002US-0411037P.
PR 17-SEP-2002; 2002US-0411041P.
PR 17-SEP-2002; 2002US-0411045P.
PR 17-SEP-2002; 2002US-0411046P.
PR 17-SEP-2002; 2002US-0411048P.
PR 17-SEP-2002; 2002US-0411052P.
PR 17-SEP-2002; 2002US-0411055P.
PR 17-SEP-2002; 2002US-0411073P.
PR 17-SEP-2002; 2002US-0411082P.
PR 17-SEP-2002; 2002US-0411101P.
PR 17-SEP-2002; 2002US-0411111P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463708P.
PR 18-APR-2003; 2003US-0463716P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467199P.
PR 02-MAY-2003; 2003US-0467201P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 22-MAY-2003; 2003US-0471336P.
PR 22-MAY-2003; 2003US-0472420P.
PR 22-MAY-2003; 2003US-0472430P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476641P.
PR 08-JUL-2003; 2003US-0485218P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485224P.
PR 14-JUL-2003; 2003US-0486446P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486891P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493341P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
PR 08-AUG-2003; 2003US-0493577P.
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX PA William LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
XX PI Halenbeck RF, Huang MM, Kothakota S, Haishan L, Linnemann T;
XX PI Pierce K, Wang Y, Wong JGP, Wu G, Zhang H;
XX WPI; 2004-348438/32.
XX New nucleic acid molecule for diagnosing, preventing or treating diseases
XX such as proliferative (e.g. cancer), inflammatory, immune, metabolic,
XX genetic, bacterial and viral diseases.
XX PT

XX Claim 1; SEQ ID NO 2020; 428pp; English.

XX The present invention relates to an isolated nucleic acid molecule

XX encoding a polypeptide which is believed to be cytostatic,

XX antiinflammatory, immunosuppressive, antibacterial and virucidal. The

XX composition and methods are useful for diagnosing, preventing and

XX treating diseases such as proliferative (e.g. cancer), inflammatory,

XX immune, metabolic, genetic, bacterial and viral diseases. The present

XX sequence represents a human secreted protein. The present sequence is

XX available on WIPWEB and is not in the specification.

SQ Sequence 185 AA;

Query Match 71.7%; Score 746; DB 8; Length 185;

Best Local Similarity 85.6%; Pred. No. 4.2e-59;

Matches 160; Conservative 9; Mismatches 12; Indels 6; Gaps 4;

QY 24 GLETGKIQDLIHLRQAYLSEHAEANEEDVLGDETEREETKPIELPVKEEPPPEKTVTD 83

DB 5 GLETGKIQDLIHLRQAYLSEHAEANEEDVLGDETEREETKPIELPVKEEPPPEKTVTD 59

QY 84 VAAEKVKVITSEIPQTERMQKRAERFNVPSLESKKAARAFGISSVPTKGLSSDNKP 143

DB 60 VAAEKVKVITSEIPQTERMQKRAEQSFVPVSLESKKAQAARFGISSVPTK-VCHLNT 118

QY 144 MVNLDKIKERAQFGLNVSSISRSKSEDEKIKRKRFGIVTSAGTGTTEDTEAKKRKR 203

DB 119 MVNLDKPKERAQFGLNVSSISRSKSEDDKIKRKRFGIVTSAGTGTTEDTEAKKRKR 178

QY 204 AERFGIA 210

DB 179 AERFGIA 185

RESULT 8

AAG00732

ID AAG00732 standard; protein; 149 AA.

AC AAG00732;

XX 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 4813.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

PN 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR N-PSDB; AAC00738.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 13; SEQ ID NO 4813; 71pp + Sequence Listing; English.

XX The present sequence is a polypeptide encoded by one of a large number of

XX 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were

CC prepared from total human RNAs or polyA+ RNAs derived from 30 different

CC tissues. EST sequences usually correspond mainly to the 3' untranslated

CC region (UTR) of the mRNA because they are often obtained from oligo-dT

CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA

CC sequences derived from the 5' ends of mRNAs and even in those cases where

CC longer cDNA sequences have been obtained, the full 5' UTR is rarely

CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can

CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs

CC are also used in diagnostic, forensic, gene therapy and chromosome

CC mapping procedures. They are used to obtain upstream regulatory sequences

CC and to design expression and secretion vectors

XX Sequence 149 AA;

Query Match 71.3%; Score 742; DB 3; Length 149;

Best Local Similarity 100.0%; Pred. No. 7.3e-59;

Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATETVELHKLKLAELKQCLARGLETGKIQDLIHLRQAYLSEHAEANEEDVLGDET 60

DB 1 MATETVELHKLKLAELKQCLARGLETGKIQDLIHLRQAYLSEHAEANEEDVLGDET 60

QY 61 ESEETKPIELPVKEEPPPEKTVDAAEKKVKVITSEIPQTERMQKRAERFNVPSLESK 120

DB 61 ESEETKPIELPVKEEPPPEKTVDAAEKKVKVITSEIPQTERMQKRAERFNVPSLESK 120

QY 121 ARAARFGISSVPTKGLSSDNKPMVNLDK 149

DB 121 ARAARFGISSVPTKGLSSDNKPMVNLDK 149

RESULT 9

ABB63720

ID ABB63720 standard; protein; 308 AA.

AC ABB63720;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 17952.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR N-PSDB; ABL07823.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.

XX Disclosure; SEQ ID NO 17952; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention


```
XX 02-OCT-2002; 2002US-0416186P.
XX (NUVE-) NUVELO INC.
XX
XX Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y;
XX Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;
XX WPI; 2004-668857/65.
XX N-PSDB; ADS09800.
XX
XX New polynucleotide, useful in preparing a composition for diagnosing or
XX treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
XX aplastic anemia or cancer for promoting wound healing.
XX
XX Claim 20; SEQ ID NO 721; 718pp; English.
XX
XX The invention relates to a novel isolated polynucleotide and the encoded
XX polypeptide. The molecules of the invention demonstrate antiinflammatory,
XX neuroprotective, antianaemic, cytostatic and vulnerary activities and may
XX be useful in preparing a composition for diagnosing or treating
XX inflammatory, haematopoietic, immune, neurodegenerative or stem cell
XX disorders, such as aplastic anaemia or cancer, as well as for promoting
XX wound healing. The molecules may also be utilised during gene therapy
XX procedures. The current sequence is that of a human therapeutic protein
XX of the invention. The current sequence is not shown explicitly within the
XX specification but can be accessed from the WIPO web-site.
XX
XX Sequence 747 AA;
Query Match 14.9%; Score 155; DB 8; Length 747;
Best Local Similarity 26.2%; Pred. No. 4.5e-05;
Matches 61; Conservative 37; Mismatches 77; Indels 58; Gaps 9;
QY 6 VELHKLKLAELKQFCIARGLETGKIKODLIHRLQAYLE-EHAE-EA----- 50
DB 1 MEVRLKLVTELSELQRRGLDSRLGKVDLAQLQALDAEMLEDEAGGGGAGPGGACKAE 60
QY 51 -----NEEDVLGDETEEEETKPIELPVKEEPEPEKTVDAEKKVVKITSEI 97
DB 61 PRVAAGGGPGGDEEB---DEEEEDDEBALLEDEDEPP-----PAQALGQAAQPPPEP 113
QY 98 POTERMOKRAERNFVPSLESKKAARAARFGISSVPTKGLSSDNKPMVNDLKLKRA--Q 155
DB 114 PEAANEAAAE---PDAEKPAAETAGSGVNGEEOGLKREE-----DEPEERSGDE 164
QY 156 RFLGNV--SSISRSKSEDEKLKKRKRFGITVTSAGTGTTEDEAKRKRAER 206
DB 165 TPGSEVPGDKAAEEQGDQDSEKSP-----AGSDGERRGVKQRDEK 207
RESULT 12
ABB61567
ID ABB61567 standard; protein; 795 AA.
XX AC ABB61567;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 11493.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL05670.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 11493; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 795 AA;
Query Match 14.9%; Score 155; DB 4; Length 795;
Best Local Similarity 21.0%; Pred. No. 4.9e-05;
Matches 55; Conservative 49; Mismatches 86; Indels 72; Gaps 8;
QY 4 ETVELHKLKLAELKQFCIARGLETGKIKODLIHRLQAYLE-EHAE-EA----- 46
DB 2 DVAKLEKMKVLDRLNELQSRGLDTGKGAVLVERLAYVEGGAGDGENAPVTPSRQRRT 61
QY 47 -----EEANEEDVLGDETEEEETKPIELPVKEEPEPEKTV 82
DB 62 RSMRSFSPVQAPVAAPAEVPLDTLEEEQEDKTVQPEPESEQPAAEPEPESEPEAE 121
QY 83 DVAAEKVVVKITSEI-----POTERMOKRAERNFVPSLESKKAARAARFGISSVPTK-- 135
DB 122 PAAAVTDDTTVAQVNEESQPEPEFDEKSETDDKQETIEEAVPAVVPQNEVADEPMEED 181
QY 136 --GLSSDNKPMVNDLKLKRAQRFGLNVSSIS--RKSEDEKLKKRKRFGITVTSAGTG 191
DB 182 HDAAPBEQEPOTTEPEVBEKPAB-----STVAEHQSGNDSQKMDVDEE-----DSAAPK 230
QY 192 TTEDTE-----AKRKRA 204
DB 231 TAETEPAAKPEDQPPERRKRS 252
RESULT 13
AAG20995
ID AAG20995 standard; protein; 185 AA.
XX AC AAG20995;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 23391.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EF1033405-A2.
XX
XX 06-SEP-2000.
XX
```


PF 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-01231180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0128246P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139751P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.

PR	28-SEP-1999;	99US-0156458P.	XX	25-FEB-2000;	2000EP-00301439.	XX
PR	29-SEP-1999;	99US-0156596P.	PR	25-FEB-1999;	99US-0121825P.	PR
PR	04-OCT-1999;	99US-0157117P.	XX	05-MAR-1999;	99US-0123180P.	PR
PR	03-OCT-1999;	99US-0157753P.	PR	09-MAR-1999;	99US-0123548P.	PR
PR	06-OCT-1999;	99US-0157865P.	PR	23-MAR-1999;	99US-0125788P.	PR
PR	07-OCT-1999;	99US-0158029P.	PR	25-MAR-1999;	99US-0126264P.	PR
PR	08-OCT-1999;	99US-0158232P.	PR	29-MAR-1999;	99US-0126785P.	PR
PR	12-OCT-1999;	99US-0158369P.	PR	01-APR-1999;	99US-0127462P.	PR
PR	13-OCT-1999;	99US-0159293P.	PR	06-APR-1999;	99US-0128234P.	PR
PR	13-OCT-1999;	99US-0159294P.	PR	08-APR-1999;	99US-0128714P.	PR
PR	14-OCT-1999;	99US-0159329P.	PR	16-APR-1999;	99US-0129845P.	PR
PR	14-OCT-1999;	99US-0159330P.	PR	19-APR-1999;	99US-0130077P.	PR
PR	14-OCT-1999;	99US-0159331P.	PR	21-APR-1999;	99US-0130449P.	PR
PR	14-OCT-1999;	99US-0159637P.	PR	23-APR-1999;	99US-0130510P.	PR
PR	18-OCT-1999;	99US-0159638P.	PR	28-APR-1999;	99US-0130891P.	PR
PR	21-OCT-1999;	99US-0159584P.	PR	30-APR-1999;	99US-0131449P.	PR
PR	21-OCT-1999;	99US-0160741P.	PR	30-APR-1999;	99US-0132048P.	PR
PR	21-OCT-1999;	99US-0160767P.	PR	04-MAY-1999;	99US-0132407P.	PR
PR	21-OCT-1999;	99US-0160768P.	PR	05-MAY-1999;	99US-0132484P.	PR
PR	21-OCT-1999;	99US-0160770P.	PR	06-MAY-1999;	99US-0132485P.	PR
PR	21-OCT-1999;	99US-0160814P.	PR	06-MAY-1999;	99US-0132486P.	PR
PR	21-OCT-1999;	99US-0160815P.	PR	07-MAY-1999;	99US-0132487P.	PR
PR	22-OCT-1999;	99US-0160980P.	PR	11-MAY-1999;	99US-0132863P.	PR
PR	22-OCT-1999;	99US-0160981P.	PR	14-MAY-1999;	99US-0134256P.	PR
PR	22-OCT-1999;	99US-0160989P.	PR	14-MAY-1999;	99US-0134218P.	PR
PR	25-OCT-1999;	99US-0161404P.	PR	14-MAY-1999;	99US-0134219P.	PR
PR	25-OCT-1999;	99US-0161405P.	PR	14-MAY-1999;	99US-0134221P.	PR
PR	25-OCT-1999;	99US-0161406P.	PR	14-MAY-1999;	99US-0134370P.	PR
PR	26-OCT-1999;	99US-0161359P.	PR	18-MAY-1999;	99US-0134768P.	PR
PR	26-OCT-1999;	99US-0161360P.	PR	19-MAY-1999;	99US-0134941P.	PR
PR	26-OCT-1999;	99US-0161361P.	PR	20-MAY-1999;	99US-0135124P.	PR
PR	28-OCT-1999;	99US-0161920P.	PR	21-MAY-1999;	99US-0135353P.	PR
PR	28-OCT-1999;	99US-0161952P.	PR	24-MAY-1999;	99US-0135629P.	PR
PR	28-OCT-1999;	99US-0161993P.	PR	25-MAY-1999;	99US-0136021P.	PR
PR	29-OCT-1999;	99US-0162142P.	PR	27-MAY-1999;	99US-0136392P.	PR
Query Match 13.4%; Score 139; DB 3; Length 185;						
Best Local Similarity 31.2%; Pred. No. 0.00021;						
Matches 55; Conservative 22; Mismatches 67; Indels 32; Gaps 7;						
Qy	57	GDETEEBETPIELPVKEEPPPEKTVD-----VAAEKKVKITSEIPOT-----	100			
Db	17	GLSTGENPKIVDLNITELDRTDDILDGEVKGFSDSGEKEETDSNGIGSTAGVDSGDIS	76			
Qy	101	-----ERMQKRAERFNPVSL--ESKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKER	153			
Db	77	PVDDIQKKIRRAERFGVSVKLTTEEKRNRAERFGTVAAAVVNGSEGTAKAEL-KRKAR	135			
Qy	154	AQRGGL-NVSSI3RKSDEDEKLKKRERFGIVTSSAGTGTTEAKRK-RAERF	207			
Db	136	ADRFVPSATSTTDKTBEEAKKARLARFGKETK-----VDSAEENKKRKAALRF	185			
RESULT 14						
AAG22070						
ID	AAG22070 standard; protein; 214 AA.					
XX	AC AAG22070;					
XX	AC AAG22070;					
DT	17-OCT-2000 (first entry)					
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 24859.					
DE	Arabidopsis thaliana					
XX	Protein identification; signal transduction pathway; metabolic pathway;					
KW	hybridisation assay; genetic mapping; gene expression control; promoter;					
KW	termination sequence.					
OS	Arabidopsis thaliana.					
XX	EP1033405-A2.					
PN	06-SEP-2000.					
XX						
PD						

PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145102P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145219P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 02-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0149930P.
PR 26-AUG-1999; 99US-0150566P.
PR 27-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.

PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 13.4%; Score 139; DB 3; Length 214;
Best Local Similarity 31.2%; Pred. No. 0.00026;
Matches 55; Conservative 22; Mismatches 67; Indels 32; Gaps 7;

Qy 57 GDETEEEETPIELPVKEEPEPEKTVD-----VAAEKVKVITSEIPQT----- 100
Db 17 GLSTGENPKKIVDLNTTDLTDDILDGEVKGFSDSGEKKEETDSNGIGSTAGVDSGDIS 76
Qy 101 -----ERMQKRAERNVPVSL--ESKKAARAARFGISSVPTKGLSSDNKPMVLDLKER 153
Db 77 PVDDIQKKIRRAERFGVSVKLTTEEKKNSRAERFGTVAAAVVNGSEGTAKBEL-KRKAR 135
Qy 154 AQRFGGL-NVSSISRKSDDEKLKKRKERFGIVTSSAGTGTTEDEAKRK-RARERF 207
Db 136 ADRFGVPSATSTTDKTEEEAKKARLARFCKETK-----VDSAEENKKARALRF 185

RESULT 15

ADRO8732
ID ADRO8732 standard; protein; 750 AA.

XX AC ADRO8732;

DT 04-NOV-2004 (first entry)

DE Human protein useful for treating neurological disease Seq 2238.

XX human; oligo-capping method; diagnostic marker; gene therapy;

KW osteoporosis; neurological disease; Alzheimer's disease;

KW Parkinson's disease; dementia; short memory; cancer;

KW sense or motor function; emotional reaction; fear response; panic;

KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;

KW tranquilliser.

XX Homo sapiens.

OS

```
XX EP1447413-A2.
PN
XX
XX
PD 18-AUG-2004.
XX
XX 12-FEB-2004; 2004EP-00003145.
XX
XX 14-FEB-2003; 2003JP-00102207.
PR
XX 09-MAY-2003; 2003JP-00131452.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
XX
DR WPI; 2004-583265/57.
DR N-PSDB; ADR06776.
XX
XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
XX Claim 1; SEQ ID NO 2338; 2686pp; English.
XX
XX This invention relates to novel, isolated full length human cDNA
CC molecules and the encoded proteins thereof. Specifically, it refers to
CC cDNA clones obtained by an oligo-capping method, where none of these
CC clones are identical to any known human mRNAs. The present invention
CC describes an immunoassay to identify agonists and antagonists, as well as
CC antibodies, antisense molecules and siRNAs that can all be used to bind
CC to and modulate expression of the cDNA molecules. As such, these
CC molecules are useful for diagnostic markers or therapeutic targets for
CC the various diseases or morbid states. In particular, they are useful in
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC disease, Parkinson's disease, dementia, short memory and various cancers,
CC as well as for maintaining equilibrium of sense or motor function, and
CC for treating emotional reaction, fear response and panic. Accordingly,
CC they exhibit osteopathic, neuroprotective, neurotropic, antiparkinsonian,
CC cytotatic and tranquiliser activities. This polypeptide is a protein
CC encoded by a full length human cDNA sequence of the invention. NOTE: This
CC sequence is not given in the sequence listing of the specification but
CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC office.
XX
XX Sequence 750 AA;
SQ
Query Match 13.3%; Score 138; DB 8; Length 750;
Best Local Similarity 25.4%; Pred. No. 0.0015;
Matches 60; Conservative 27; Mismatches 77; Indels 72; Gaps 9;
QY 1 MATETVELHKLKLAELKQECLEGLTKGKQDLIHLQVLEEH-----AEEAN--- 51
Db 1 MSSSPNVNKKLVSELKELKKRLSDGKLKAEUWELRQAALDDDEAGRPAPWPGNGDD 60
QY 52 -----EDVLGDETE-----EETKPIELPVKEEPEPEKTVDVAAE-----KK 89
Db 61 QGFQEGEDLGDDEEGAGDENGHEGQPP---PATQQQPQQQGAKEAGKSSGPTS 117
QY 90 VVKITSIPOTERNQKKAERPNVPVLSLESKKAARAFGISVPTKGLSSDNKPMVNLDK 149
Db 118 LFAVTVAPPGARQGOQQAQ-----GKKKAGGGGGGRPGAPAG---DGGK----- 159
QY 150 LKERAQRFGLNVLSIRKSEDDDEKLKKRERFGIVTSSAGTGTTEAKKRAE 205
Db 160 ---TEQGGDKKRGVKRPREDH-----GRGYFEYIENKYSRAK 195
```

Search completed: October 5, 2005, 05:57:27
Job time : 108 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 05:59:14 ; Search time 1324 Seconds
(without alignments)
65.842 Million cell updates/sec

Title: US-09-788-476A-2
Perfect score: 1040
Sequence: 1 MATETVELHLKLAELKQEC.....GTTEDTEAKRKRAERFGIA 210

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues

Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1040	100.0	210	9 US-09-788-476A-2	Sequence 2, Appli
2	1040	100.0	228	9 US-09-764-846-218	Sequence 218, App
3	1040	100.0	228	14 US-10-091-483-218	Sequence 218, App
4	222	21.3	308	20 US-11-097-143-17952	Sequence 17952, A
5	160	15.4	169	16 US-10-425-115-335173	Sequence 335173,
6	155	14.9	734	16 US-10-483-506-6	Sequence 6, Appli
7	155	14.9	795	20 US-11-097-143-11493	Sequence 11493, A
8	146.5	14.1	165	15 US-10-424-599-156346	Sequence 156346,
9	146	14.0	115	15 US-10-425-114-48886	Sequence 48886, A
10	143	13.8	175	16 US-10-437-963-176147	Sequence 176147,
11	140	13.5	153	16 US-10-767-701-32849	Sequence 32849, A

12	138.5	13.3	189	15	US-10-424-599-156348	Sequence 156348,
13	138.5	13.3	212	15	US-10-425-114-44717	Sequence 44717, A
14	137.5	13.2	540	14	US-10-106-698-5766	Sequence 5766, Ap
15	137.5	13.2	1150	15	US-10-433-544-4	Sequence 4, Appli
16	136.5	13.1	1150	14	US-10-316-532-4	Sequence 4, Appli
17	136.5	13.1	1172	17	US-10-828-815-35	Sequence 35, Appli
18	133	12.8	1146	14	US-10-316-532-2	Sequence 2, Appli
19	133	12.8	1153	17	US-10-828-815-33	Sequence 33, Appli
20	124	11.9	775	10	US-09-934-455-450	Sequence 450, App
21	122.5	11.8	805	15	US-10-369-493-152	Sequence 152, App
22	122	11.7	2274	20	US-11-097-143-2763	Sequence 2763, Ap
23	120.5	11.6	450	15	US-10-094-749-2630	Sequence 2630, Ap
24	120.5	11.6	824	20	US-11-009-554-23	Sequence 23, Appli
25	120.5	11.6	2020	15	US-10-369-493-5128	Sequence 5128, Ap
26	120.5	11.6	2020	15	US-10-369-493-5129	Sequence 5129, Ap
27	119.5	11.5	686	16	US-10-425-115-213836	Sequence 213836,
28	119.5	11.5	1332	9	US-09-982-091A-4	Sequence 4, Appli
29	119	11.4	268	15	US-10-412-699B-730	Sequence 730, App
30	119	11.4	811	15	US-10-264-049-3051	Sequence 3051, Ap
31	118.5	11.4	845	15	US-10-205-331-61	Sequence 61, Appli
32	118	11.3	303	16	US-10-425-115-213843	Sequence 213843,
33	116.5	11.2	36946	18	US-10-840-512-155	Sequence 155, App
34	116	11.2	713	16	US-10-437-963-177006	Sequence 177006,
35	116	11.2	806	16	US-10-684-422-184	Sequence 184, App
36	116	11.2	806	20	US-11-009-554-22	Sequence 22, Appli
37	115.5	11.1	856	18	US-10-764-425-154	Sequence 154, App
38	115.5	11.1	1150	20	US-11-097-143-4179	Sequence 4179, Ap
39	115	11.1	428	14	US-10-254-995-7	Sequence 7, Appli
40	115	11.1	428	20	US-11-062-080-7	Sequence 7, Appli
41	115	11.1	439	10	US-09-056-019-37	Sequence 37, Appli
42	115	11.1	439	16	US-10-751-702-37	Sequence 37, Appli
43	115	11.1	439	20	US-11-032-644-37	Sequence 37, Appli
44	115	11.1	1359	17	US-10-732-923-8707	Sequence 8707, Ap
45	115	11.1	1359	17	US-10-732-923-8708	Sequence 8708, Ap

ALIGNMENTS

RESULT 1

US-09-788-476A-2
; Sequence 2, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788,476A
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 210
; TYPE: PRT
; ORGANISM: human
US-09-788-476A-2

Query Match 100.0%; Score 1040; DB 9; Length 210;
Best Local Similarity 100.0%; Pred. No. 2.9e-76;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MATETVELHLKLAELKQEC	LARGLETKGIKQDLIHLRQAYLEHAEANEEDVLGDET	60
Db	1	MATETVELHLKLAELKQEC	LARGLETKGIKQDLIHLRQAYLEHAEANEEDVLGDET	60
Qy	61	EEETKPIELPVKEEPEPKTV	VAACKVKVKTITSETPQTERMQKRAERFNVPSLESKK	120
Db	61	EEETKPIELPVKEEPEPKTV	VAACKVKVKTITSETPQTERMQKRAERFNVPSLESKK	120
Qy	121	AAARAARFGISVPPTKGLSS	DNKPMVNDKUKERAQRLGNVSSISRSKSEDEKLUKKRER	180
Db	121	AAARAARFGISVPPTKGLSS	DNKPMVNDKUKERAQRLGNVSSISRSKSEDEKLUKKRER	180

```
Qy 181 FGIWTSAGTGTEDTEAKKRAERFGIA 210
|||||
Db 181 FGIWTSAGTGTEDTEAKKRAERFGIA 210
|||||

RESULT 2
US-09-764-846-218
; Sequence 218, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ12
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 218
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-846-218

Query Match 100.0%; Score 1040; DB 9; Length 228;
Best Local Similarity 100.0%; Pred. No. 3.2e-76;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAFEEANEEDVLGDET 60
|||||
Db 19 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAFEEANEEDVLGDET 78
|||||

Qy 61 EEBETKPIELPVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERNVPVSLESKK 120
|||||
Db 79 EEBETKPIELPVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERNVPVSLESKK 138
|||||

Qy 121 AAAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
|||||
Db 139 AAAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 198
|||||

Qy 181 FGIWTSAGTGTEDTEAKKRAERFGIA 210
|||||
Db 199 FGIWTSAGTGTEDTEAKKRAERFGIA 228
|||||

RESULT 3
US-10-091-483-218
; Sequence 218, Application US/10091483
; Publication No. US20030049650A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ12C1
; CURRENT APPLICATION NUMBER: US/10/091,483
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 348
; Prior application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 218
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-091-483-218

Query Match 100.0%; Score 1040; DB 14; Length 228;
Best Local Similarity 100.0%; Pred. No. 3.2e-76;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAFEEANEEDVLGDET 60
|||||
Db 19 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAFEEANEEDVLGDET 78
|||||

Qy 61 EEBETKPIELPVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERNVPVSLESKK 120
|||||
```

```
Db 79 EEBETKPIELPVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERNVPVSLESKK 138
|||||
Qy 121 AAAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
|||||
Db 139 AAAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 198
|||||
Qy 181 FGIWTSAGTGTEDTEAKKRAERFGIA 210
|||||
Db 199 FGIWTSAGTGTEDTEAKKRAERFGIA 228
|||||

RESULT 4
US-11-097-143-17952
; Sequence 17952, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17952
; LENGTH: 308
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-17952

Query Match 21.3%; Score 222; DB 20; Length 308;
Best Local Similarity 27.3%; Pred. No. 7.7e-10;
Matches 83; Conservative 40; Mismatches 77; Indels 104; Gaps 10;

Qy 7 ELHKLKLAELKQECCLARGLETGKIQDLIHRLOA-----VLEHAEAEANEED----- 54
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 5 DVTMKVADLKRELKRLGLAVNGKTELQDLRLOTALLEGDSLEDSAIADAIDDDVVSFT 64
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 55 -----VLGDETEEE-----ETKPIEL----- 70
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 65 DEDEHKLDGDNDELLKSPVSTPTTVAIPDLIAEETSSAPDAAAPTKKIVLKRNNSSQ 124
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 71 -----PVKEEPEP-----EKTVDV-----AAEKVVVKITS 95
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 125 STGTVASTGTTTSPKNEAPAAAASDSTGETPTKKHKPIVVGPKTEGKPSGDKKLNLTA 184
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 96 EIPQTERMQKRAERNVPVPSLESKKAARAAREFGISSVPTKGLSSDNKPMVNLDK--LKE 152
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 185 Q-----ERLEIRAKKFGITPPAVANTATAVAINKSSASITANKGNKETEEQKEALKK 240
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 153 RQRFGLNVSSISRKSEDEKLKKRERFGIIVTSAGTGTEDTEA-----KKRKAER 206
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 241 RAERFGVVPDKAPTSKADRLQKKRERFGAGAVSAATTTPTTTTESKDAWSEKARARLER 300
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```


; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11493
; LENGTH: 795
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-11493

Query Match 14.9%; Score 155; DB 20; Length 795;
Best Local Similarity 21.0%; Pred. No. 0.00065;
Matches 55; Conservative 49; Mismatches 86; Indels 72; Gaps 8;

QY 4 ETVELHKLAEKQELRGLETGKIKODLIHRLQAYLE-----EHA----- 46
DB 2 DVAKLEKMKVVDLRLNELQSLGLTKGVKVLVERLRAVVEGGAGDGENAPVTPSRQRRT 61
QY 47 -----EEANEEDVLGDTEEEETKPIELPVKEPEPEKTV 82
DB 62 RMSRSPSPVQAAAPVAAPVLDLLEEEQEDTKVPQPEPESEQPAAPPEPEQSEPEEAE 121
QY 83 DVAAEKVKVITSEI-----PQTERMQKRAERFNVPSLESKKAARAARFGISVPTK-- 135
DB 122 PAAAVTEDTTVNOAVNSESQPEPEFEDEKSETDDKQETIEEAVPAVVPQNEVADEPMEED 181
QY 136 --GLSSDNKPMVLDLKLKERAQFGLNVSSIS--RKSEDDKLLKRAERFGIVTSSAGTG 191
DB 182 HDAAPQEQTQTEEPVEEKPAE-----STVAEHQNSGDSQKVDVDEE-----DSAAPK 230
QY 192 TTEDTE-----AKKRKA 204
DB 231 TABETEPAAKPEQPQPPERRKRS 252

RESULT 8
US-10-424-599-156346
; Sequence 156346, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156346
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112200C.1.pap
US-10-424-599-156346

Query Match 14.1%; Score 146.5; DB 15; Length 165;
Best Local Similarity 31.1%; Pred. No. 0.00046;
Matches 42; Conservative 26; Mismatches 46; Indels 21; Gaps 5;

QY 75 EEPPEKTVDAEKKVKVITSEIPQTERMQKRAERFNVPSLESKKAARAARFGISSVPT 134

DB 13 QENPNKTLD-PTPEPDPDIPATEDATK-----GSDAKDPADADADAITSPPP 58
QY 135 KGLSSDNKPMVLDLKLKERAQFGLNVSSISRKSEDDKLLKRAERFGIVTSSAGTGTTT 194
DB 59 D--SGNDAPLSDIQKMRRAERFGISV-QLSEK-----EKRNSRAERFGTVSASQSEPSK 111
QY 195 DTEAKKRKAERFGI 209
DB 112 SEDLKKRAERFGM 126

RESULT 9
US-10-425-114-48886
; Sequence 48886, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48886
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3912-045-B5_FLI.pap
US-10-425-114-48886

Query Match 14.0%; Score 146; DB 15; Length 115;
Best Local Similarity 40.7%; Pred. No. 0.00033;
Matches 44; Conservative 12; Mismatches 32; Indels 20; Gaps 5;

QY 105 KRAERFNVFPV--SLSEKKAARAARFGISSVPTKGLSSDNKPMVLDLKLKERAQFGLNV 162
DB 1 RRAERFGMPVLMSEEEKSRRAERFGTSSSVK-----EEBKKSRAERFGLASP 50
QY 163 SISRKSEDDKLLKRAERFGIVTSSAGTGTTEDTEAKKRKAERFGIA 210
DB 51 S---SSDEEAKKARLERFG---QSANVDKAE--EEKKARAARFAEA 90

RESULT 10
US-10-437-963-176147
; Sequence 176147, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 176147
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Oryza sativa


```
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_73924C.1.pap
US-10-437-963-176147

Query Match      13.8%; Score 143; DB 16; Length 175;
Best Local Similarity 39.4%; Pred. No. 0.00095;
Matches 43; Conservative 14; Mismatches 32; Indels 20; Gaps 5;

Qy 101 ERMQKRAERFN--VPVSLESKAAARAFGISSVPTKGLSSDNKPMVNLDKLKERARFG 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 EKMRRARERFCTAVVMSEERSSRAERFGTG-----SSNEK--ABEQKRSRAERFG 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 159 LNVSSISRKSDDEKLKKRKRFQIVTSSAGTGTTEDTEAKRKRAERF 207
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 LASSS----AEDAKKARLERFG-----QSTNVDKGEERKARLARF 148
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-10-767-701-32849
; Sequence 32849, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 32849
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(153)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C19564_1.pap
US-10-767-701-32849

Query Match      13.5%; Score 140; DB 16; Length 153;
Best Local Similarity 38.1%; Pred. No. 0.0014;
Matches 40; Conservative 12; Mismatches 31; Indels 22; Gaps 4;

Qy 101 ERMQKRAERFNPV--SLESKAAARAFGISSVPTKGLSSDNKPMVNLDKLKERARFG 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 EKMRRARERFCTPVVMSEERSSRAERFGTGSSSVK-----EEEKKRSRAEKIG 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 159 LNVSSISRKSDDEKLKKRKRFQIVTSSAGTGTTEDTEAKRK 203
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 103 LASSS----SSDEAKKARLERF-----QGTNGDKAKKKKK 137
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-10-424-599-156348
; Sequence 156348, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156348
; LENGTH: 189
```

```
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112202C.1.pap
US-10-424-599-156348

Query Match      13.3%; Score 138.5; DB 15; Length 189;
Best Local Similarity 30.7%; Pred. No. 0.0024;
Matches 46; Conservative 20; Mismatches 53; Indels 31; Gaps 5;

Qy 63 EETKPIELPVKEPEPEKTVDVAAEKVKVITSEIPQTERMQKRAERFNPVPSLESKAA 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 DPEPDIPATEDDAADTKDSAK-----NSPPPDGSGN-----DAPLSDIQKKMR 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 123 RAARFGISSVPTKGLSSDNKPMVNLDKLKERARFGLNVSSISRKSEDDDKLKK--RKER 180
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 RAERFGIS-----VOLSEKERNRAERFGTGSASQGSSEPSKSELKRAER 115
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 181 FGIVTSSAGTGTTEDTEAKRKRAERFGIA 210
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 FGMPSPPT---TTSDEEAKKKARLARFAPA 141
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-10-425-114-44717
; Sequence 44717, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 44717
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700751810_FLI.pap
US-10-425-114-44717

Query Match      13.3%; Score 138.5; DB 15; Length 212;
Best Local Similarity 30.7%; Pred. No. 0.0028;
Matches 46; Conservative 20; Mismatches 53; Indels 31; Gaps 5;

Qy 63 EETKPIELPVKEPEPEKTVDVAAEKVKVITSEIPQTERMQKRAERFNPVPSLESKAA 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 44 DPEPDIPATEDDAADTKDSAK-----NSPPPDGSGN-----DAPLSDIQKKMR 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 123 RAARFGISSVPTKGLSSDNKPMVNLDKLKERARFGLNVSSISRKSEDDDKLKK--RKER 180
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 RAERFGIS-----VOLSEKERNRAERFGTGSASQGSSEPSKSELKRAER 138
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 181 FGIVTSSAGTGTTEDTEAKRKRAERFGIA 210
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 FGMPSPPT---TTSDEEAKKKARLARFAPA 164
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-10-106-698-5766
; Sequence 5766, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 02:29:12 ; Search time 26 Seconds
(without alignments)
777.135 Million cell updates/sec

Title: US-09-788-476A-2
Perfect score: 1040
Sequence: 1 MATETVELHLKLAELKQEC.....GTTEDTEAKKRAERFGIA 210

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1040	100.0	210	2 JC7830	cytokine-inducible
2	222	21.3	245	2 T41291	conserved hypothet
3	148	14.2	344	2 T48827	hypothetical prote
4	139	13.4	214	2 T48298	hypothetical prote
5	139	13.4	1222	2 T22490	hypothetical prote
6	135.5	13.0	3488	2 T34418	hypothetical prote
7	133.5	12.8	633	2 T05005	hypothetical prote
8	125.5	12.1	218	2 S50566	hypothetical prote
9	125	12.0	1110	2 I31116	NP-180 - sea lampr
10	123.5	11.9	849	2 S00030	neurofilament trip
11	123	11.8	363	2 F87754	protein C43E11.1 l
12	122.5	11.8	399	2 E71169	hypothetical prote
13	122.5	11.8	805	2 E70474	translation initia
14	120.5	11.6	2020	2 T21174	hypothetical prote
15	119.5	11.5	1087	2 T30330	gelsolin-related p
16	119.5	11.5	1224	2 F96795	hypothetical prote
17	119	11.4	415	2 T32490	hypothetical prote
18	118.5	11.4	517	1 A39038	l-caldesmon, nonmu
19	118.5	11.4	845	2 A45669	neurofilament trip
20	117	11.2	276	2 A25345	troponin T, cardia
21	116.5	11.2	263	2 C31957	troponin T, skelet
22	116.5	11.2	644	2 S5395	neurofilament prot
23	116	11.2	318	2 C84651	hypothetical prote
24	116	11.2	806	2 S22765	heterogeneous ribo
25	115.5	11.1	856	2 T13159	E1B-55kDa-associat
26	115.5	11.1	1877	2 T21861	hypothetical prote
27	115	11.1	1359	2 T34036	hypothetical prote
28	115	11.1	7962	2 T38346	elastic titin - hu
29	114.5	11.0	763	2 T08929	hypothetical prote

30	114	11.0	325	2 T18283	hypothetical prote
31	114	11.0	419	2 G75062	probable flagella-
32	113.5	10.9	573	2 S6710	probable membrane
33	113.5	10.9	683	2 E86358	Fl2K8.13 protein -
34	113.5	10.9	720	2 T26819	hypothetical prote
35	112.5	10.8	1408	2 T45039	hypothetical prote
36	111.5	10.7	718	2 AB1258	hypothetical prote
37	110.5	10.6	375	2 S26059	probable transform
38	110.5	10.6	407	2 S23325	M2 protein precurs
39	110.5	10.6	2017	1 A36014	myosin heavy chain
40	110.5	10.6	2057	2 S61477	myosin II heavy ch
41	110	10.6	771	1 A33430	h-caldesmon - chic
42	110	10.6	1603	2 S17983	gene posterior sex
43	110	10.6	1701	2 T09127	probable erythrocy
44	109.5	10.5	1156	2 B70356	chromosome assembl
45	109	10.5	278	1 TPHTWT	troponin T, slow s

ALIGNMENTS

RESULT 1

JC7830
cytokine-inducible 29K protein, CIP29 - human
C:Species: Homo sapiens (man)
C:Date: 09-Dec-2002 #sequence_revision 09-Dec-2002 #text_change 09-Jul-2004
C:Accession: JC7830; PC7188
R:Fukuda, S.; Wu, D.W.; Stark, K.; Pelus, L.M.
Biochem. Biophys. Res. Commun. 292, 593-600, 2002
A:Title: Cloning and characterization of a proliferation-associated cytokine-inducible p
A:Reference number: JC7830; MUID:21920340; PMID:11922608
A:Accession: JC7830
A:Molecule type: DNA
A:Residues: 1-210 <FUK>
A:Cross-references: UNIPROT:P82979; GB:bankit447886
A:Accession: PC7188
A:Molecule type: protein
A:Residues: 121-137;152-168 <FU2>
C:Comment: This protein that is a novel erythropoietin-induced protein is involved in no
either phosphorylated or glycosylated, and is involved in DNA transcription. This prote
C:Genetics:
A:Gene: cip29
A:Map position: 7; 8; 12

Query Match		100.0%;	Score 1040;	DB 2;	Length 210;
Best Local Similarity		100.0%;	Pred. No. 4.4e-57;		
Matches 210;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MATETVELHLKLAELKQEC	LARGLETGKIQD	LHRLQAYLEHAE	EENEEDVLGDET 60
Db	1	MATETVELHLKLAELKQEC	LARGLETGKIQD	LHRLQAYLEHAE	EENEEDVLGDET 60
Qy	61	EEETKPIELPVKEEPE	PKTVDAEKKVKIT	SEIPQTERMOKRAER	FNPVSLSKK 120
Db	61	EEETKPIELPVKEEPE	PKTVDAEKKVKIT	SEIPQTERMOKRAER	FNPVSLSKK 120
Qy	121	AAARAARFGISSVP	TKGLSSDNKPMV	NLDKLKERAQR	FGLNVSSISRSKSEDDKLKKRKR 180
Db	121	AAARAARFGISSVP	TKGLSSDNKPMV	NLDKLKERAQR	FGLNVSSISRSKSEDDKLKKRKR 180
Qy	181	FGIVTSSAGTGT	TEDTEAKKRAER	FGIA 210	
Db	181	FGIVTSSAGTGT	TEDTEAKKRAER	FGIA 210	

RESULT 2

T41291
conserved hypothetical protein SPC31H12.03c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T41291
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, September 1998

A;Reference number: Z21985
A;Accession: T41291
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-245 <WOO>
A;Cross-references: UNIPROT:O74871; EMBL:AL031824; PIDN:CAA21220.1; GSPDB:GN00068; SPDB:
A;Experimental source: strain 972h-; cosmid c31H12
C;Genetics:
A;Gene: SPDB:SPC31H12.03C
A;Map position: 3

Query Match 21.3%; Score 222; DB 2; Length 245;
Best Local Similarity 29.4%; Pred. No. 5.9e-07;
Matches 74; Conservative 30; Mismatches 80; Indels 68; Gaps 8;

QY 11 LKLAELKQECIARGLTKG1KQDILHRLQAYLEHAEAEANEE----- 53
DB 7 LKVAELREKLAELKGLTAGNKAELVSLTAATESNDENTSNNTATDGLDAPPEDDIDWG 66
QY 54 DVLGDETEEBETKPIELPVKEEBPEKTVDAAEKK--VVKITSEIPQT----- 100
DB 67 DMENDTISTDVKNPAPESKETSAPAAAVEIEKENESIISKTSQAPETSTGAEEHQETT 126
QY 101 -----ERMOKRAERFNVPSLES--KKAARAARFGISSVPTKGLSS 139
DB 127 EESQSVSNVSSPDVAKEQEKLQRAKRGIPVDDQIKKARAARFGIQ-----QPLAS 182
QY 140 DNKPMVNDLKKERAQFGLNVSSISRKSEDEKLLKKRERFGIVTSSAGTGTTEDEAK 199
DB 183 SNNKHNNQSNPNQRS-----NSRSQRKNKNAPPSAKSKSNIL-----DDPIEA 230
QY 200 K-RKRAERFGIA 210
DB 231 KARKRAERFGVA 242

RESULT 3
T48827
hypothetical protein 68B2.110 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
A;Accession: T48827
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24341
A;Accession: T48827
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-344 <SCH>
A;Cross-references: UNIPROT:Q9F6W5; EMBL:AL353821; GSPDB:GN00112; NCSP:68B2.110
C;Genetics:
A;Gene: NCSP:68B2.110
A;Map position: 2
A;Introns: 50/1

Query Match 14.2%; Score 148; DB 2; Length 344;
Best Local Similarity 27.2%; Pred. No. 0.029;
Matches 62; Conservative 27; Mismatches 95; Indels 44; Gaps 8;

QY 7 ELHKLKLAELKQECIARGLTKG1KQDILHRLQAYLEHAEAEANEDVL-----GDE 59
DB 3 DYNMVKVPFELKLLNERSLPOTGNKADLIARLQ-----EHDQQAQKPADAAPATTKDGEA 58
QY 60 TEEBETKPIELPVKEEBPEKTVDAAEKKVVKITSEIPQTERMQKRAERFNVPSLESK 119
DB 59 EDEIDYEDDDFPAGDKKAATETTD--EKAPASAAAEATATESEPK-----APVESEAQ 109
QY 120 KAARAARFGISSVPTKGLSSDNKPMVNL-----DKLKRAQR-----FGLNVS 162
DB 110 KTTTEAAPAAAATQPAASQAQGEVAAKDTTTTATTATTAEDPPSKAEQKPAEPLFSQHL 169
QY 163 SISRKSEDEKLLKKRERFGIVTSSAGTGTTEDEAKRKRARFGIA 210

A;Reference number: Z21985
A;Accession: T41291
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-245 <WOO>
A;Cross-references: UNIPROT:O74871; EMBL:AL031824; PIDN:CAA21220.1; GSPDB:GN00068; SPDB:
A;Experimental source: strain 972h-; cosmid c31H12
C;Genetics:
A;Gene: SPDB:SPC31H12.03C
A;Map position: 3

Query Match 21.3%; Score 222; DB 2; Length 245;
Best Local Similarity 29.4%; Pred. No. 5.9e-07;
Matches 74; Conservative 30; Mismatches 80; Indels 68; Gaps 8;

QY 11 LKLAELKQECIARGLTKG1KQDILHRLQAYLEHAEAEANEE----- 53
DB 7 LKVAELREKLAELKGLTAGNKAELVSLTAATESNDENTSNNTATDGLDAPPEDDIDWG 66
QY 54 DVLGDETEEBETKPIELPVKEEBPEKTVDAAEKK--VVKITSEIPQT----- 100
DB 67 DMENDTISTDVKNPAPESKETSAPAAAVEIEKENESIISKTSQAPETSTGAEEHQETT 126
QY 101 -----ERMOKRAERFNVPSLES--KKAARAARFGISSVPTKGLSS 139
DB 127 EESQSVSNVSSPDVAKEQEKLQRAKRGIPVDDQIKKARAARFGIQ-----QPLAS 182
QY 140 DNKPMVNDLKKERAQFGLNVSSISRKSEDEKLLKKRERFGIVTSSAGTGTTEDEAK 199
DB 183 SNNKHNNQSNPNQRS-----NSRSQRKNKNAPPSAKSKSNIL-----DDPIEA 230
QY 200 K-RKRAERFGIA 210
DB 231 KARKRAERFGVA 242

RESULT 3
T48827
hypothetical protein 68B2.110 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
A;Accession: T48827
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24341
A;Accession: T48827
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-344 <SCH>
A;Cross-references: UNIPROT:Q9F6W5; EMBL:AL353821; GSPDB:GN00112; NCSP:68B2.110
C;Genetics:
A;Gene: NCSP:68B2.110
A;Map position: 2
A;Introns: 50/1

Query Match 14.2%; Score 148; DB 2; Length 344;
Best Local Similarity 27.2%; Pred. No. 0.029;
Matches 62; Conservative 27; Mismatches 95; Indels 44; Gaps 8;

QY 7 ELHKLKLAELKQECIARGLTKG1KQDILHRLQAYLEHAEAEANEDVL-----GDE 59
DB 3 DYNMVKVPFELKLLNERSLPOTGNKADLIARLQ-----EHDQQAQKPADAAPATTKDGEA 58
QY 60 TEEBETKPIELPVKEEBPEKTVDAAEKKVVKITSEIPQTERMQKRAERFNVPSLESK 119
DB 59 EDEIDYEDDDFPAGDKKAATETTD--EKAPASAAAEATATESEPK-----APVESEAQ 109
QY 120 KAARAARFGISSVPTKGLSSDNKPMVNL-----DKLKRAQR-----FGLNVS 162
DB 110 KTTTEAAPAAAATQPAASQAQGEVAAKDTTTTATTATTAEDPPSKAEQKPAEPLFSQHL 169
QY 163 SISRKSEDEKLLKKRERFGIVTSSAGTGTTEDEAKRKRARFGIA 210

DB 170 PTDAKSE-AEKRAARAARFGITTDE-----KSEEAQKARAARFGIA 210

RESULT 4
T48298
hypothetical protein F9G14.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
A;Accession: T48298
R;Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De t
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24491
A;Accession: T48298
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-214 <BEV>
A;Cross-references: UNIPROT:O9LZ08; EMBL:AL162973
A;Experimental source: cultivar Columbia; BAC clone F9G14
C;Genetics:
A;Map position: 5
A;Introns: 109/2; 138/2; 184/2; 205/3
A;Note: F9G14.80

Query Match 13.4%; Score 139; DB 2; Length 214;
Best Local Similarity 31.2%; Pred. No. 0.062;
Matches 55; Conservative 22; Mismatches 67; Indels 32; Gaps 7;

QY 57 GDETEEEETKPIELPVKEEBPEKTVD-----VAAEKKVVKITSEIPQT----- 100
DB 17 GLSTGENPKIVDLNLTTELDRTDILDCGVKGFSDGEGKEETDSNGIGTAGVDSGDLS 76
QY 101 -----ERMOKRAERFNVPSL--ESKKAARAARFGISSVPTKGLSSDNKPMVNDLKKER 153
DB 77 PVDDIQKIRRAERFGVSVKLTTEEKRNRAERFGTVAAAVVNGSEGTAKAEL--KRKAR 135
QY 154 AQRFGI-LNVSSISRKSEDEKLLKKRERFGIVTSSAGTGTTEDEAKKPK-RAERF 207
DB 136 ADRFGVFSATSTDTKTEEEAKKARLARGKTK-----VDSAEENKRRKARALRF 185

RESULT 5
T22490
hypothetical protein Y37A1B.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
A;Accession: T22490; T26618
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19569
A;Accession: T22490
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1222 <WIL>
A;Cross-references: UNIPROT:Q9XTH8; EMBL:Z822268; PIDN:CAB05201.1; GSPDB:GN00022; CESP:X3;
A;Experimental source: clone F52B11
R;McMurray, A.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z20245
A;Accession: T26618
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1222 <W12>
A;Cross-references: EMBL:AL023835; PIDN:CAA19496.1; GSPDB:GN00022; CESP:X37A1B.1
A;Experimental source: clone Y37A1B
C;Genetics:
A;Gene: CESP:Y37A1B.1
A;Map position: 4
A;Introns: 78/3; 127/2; 165/3; 450/1; 500/2; 620/3; 728/3; 1080/1; 1157/3

Query Match 13.4%; Score 139; DB 2; Length 1222;
Best Local Similarity 27.3%; Pred. No. 0.38;

```

Matches 60, Conservative 34; Mismatches 82; Indels 44; Gaps 9
Qy 11 LKLAELKQECIARGLETGKIGKODLIHRLQAYLEEHAEFEANEEDVLGDTEEEETKPIEL 70
Db 663 MKVAELRVELELRGLETGKIGKTLVLQRLQALD--TEKAAEASVAARDVEMRDA--AEN 717
Qy 71 PVKEEPEPEK-----TVDVAAEKVKVKTSETIPQTERMQKRAERNFNPVSVLESKKAAR 123
Db 718 AVKQEGEENPAAFIAPSIBETTKAKTBAEAKKEAEAKRKKKEQGLE-----KEKKKKR 772
Qy 124 AARFGISSVPTKGLSSDNKPMWNLDKLKERARFG---LNVSSISRSKSEDEK----- 173
Db 773 EALEKHQYLP-----KDKKILVFPFK-SFKSGKGFCDKVLSSLLDVRHDDNKENQPEVS 826
Qy 174 -----LKKRKRERFGIVTSSAGTGTTEDEAKRKRKAE 205
Db 827 LFAEAFKEMIERNAAFIYETLANCG--DRDAEKKRDE 863

RESULT 6
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: Z21521
A:Accession: T34418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-references: EMBL:U80022; PIDN:AA235885.1; GSPDB:GN00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
C:Genetics:
A:Gene: CESP:F12F3.3
A:Map position: 5
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 13.0%; Score 135.5; DB 2; Length 3488;
Best Local Similarity 23.6%; Pred. No. 1.9;
Matches 56; Conservative 43; Mismatches 77; Indels 61; Gaps 8;

Qy 2 ATETVELHLKLAELKQECIARGLETGKIGKODLIHRLQAYLEEHAEFEANEEDVLGDTE 61
Db 1361 ATKVADAEKLEKQKKAABIAIEIQKEQEKLAQESRLEDEAKKSA-EKQKLESETK 1419
Qy 62 BEETKIPLPVKEPEPEKTYDVAAEKKVKVITSEIPQTERMQK-----RAERNV 112
Db 1420 SKQT-----EAPKESVDEKPKKKVLKKTEKSDSSISQKSAKSTVDAATLES 1470
Qy 113 PVSLESKKAARAARFGISSVPTKGLSSDNKPMWNLDKLKERARFGNLSVSRKSEDE 172
Db 1471 DFNLVEKTKVQK----VEQSPEDESTATIK-----RDPAQK-----TEEISKQDDGDE 1514
Qy 173 K-----LKKRKRERFGIVTSSAG-----TGTTEDETAKKRKR 203
Db 1515 KTTTDTGKPPPESEATPKGRVVKTKQKSDSVASDASLADVKLSDDVEEKPKKK 1571

RESULT 7
T05005
hypothetical protein Tl9P19.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T05005
R:Bevan, M.; Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Hoheisel, J.; Mewes
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15394
A:Accession: T05005
A:Molecule type: DNA
A:Residues: 1-633 <BEV>

```

A/Cross-references: UNIPROT:O65655; EMBL:AL022605
A/Experimental source: cultivar Columbia; BAC clone T19P19
C/Genetics:
A/Map position: 4
A/Introns: 385/1; 448/1; 498/3
A/Note: T19P19.70

Query Match 12.8%; Score 133.5; DB 2; Length 633;
Best Local Similarity 25.2%; Pred.No.0.42;
Matches 59; Conservative 34; Mismatches 66; Indels 75; Gaps 12;

Qy 8 LHKLELAEKOECLARGLETGKI KODLIHRLQAYL---EEHAE-----EEANEED 54
 : | | : ||| : | | : | | : | | : | | : | | : | | :
Db 14 IDKWVTELKEELKRRRLTTTGLKEELVRLDEALRAEQEESERINSATVAAAEKANQEP 73
 : | | : ||| : | | : | | : | | : | | : | | : | | :
Qy 55 -----VLGDETEEBETKPIE-----LPVKEEPEPKTVDAAEKKVKITSEI P QTER 102
 : | | : ||| : | | : | | : | | : | | : | | : | | :
Db 74 QMPFVTVGDRNQTTPTVVEAAFSTETTPTVAEKTPEPT-----QTKITTE----- 119
 : | | : ||| : | | : | | : | | : | | : | | : | | :
Qy 103 MQKAERFNVPVSLESKAAARAARFGISSVPTKGLSSDNKPWNLDKLKERAQRFGINVS 162
 : | | : ||| : | | : | | : | | : | | : | | : | | :
Db 120 ASAGVETTPAPVFSEPE-----VNANPV---FASDEDEKEKVDVDRDIA---GLDSS 164
 : | | : ||| : | | : | | : | | : | | : | | : | | :
Qy 163 SISR-----KXEDDEKLKRKERFGIVTSAGTGTFEDTEAKRKKA 204
 : | | : ||| : | | : | | : | | : | | : | | : | | :
Db 165 VWARDAAVVQVASSEHKSENNEPFS-----GL---DGSDSKAQPSAEVLEKSA 209
 : | | : ||| : | | : | | : | | : | | : | | : | | :

RESULT 8
S50566
hypothetical protein YER063w - yeast (*Saccharomyces cerevisiae*)
C/Species: *Saccharomyces cerevisiae*
C/Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 09-Jul-2004
C/Accession: S50566
R/Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A/Description: The sequence of *S. cerevisiae* lambda clones 6592, 4678, and 3612.
A/Reference number: S50427
A/Accession: S50566
A/Molecule type: DNA
A/Residues: 1-218 <DIE>
A/Cross-references: UNIPROT:P40040; EMBL:UI8813; NID:g1381127; PID:g603299; GSPDB:GN000
C/Genetics:
A/Gene: SGD:T01; MIPS:YER063w
A/Cross-references: SGD:S0000865
A/Map position: 5R

Query Match	12.1%;	Score 125.5;	DB 2;	Length 218;
Best Local Similarity	22.0%;	Pred. No. 0.42;		
Matches	52;	Conservative	38;	Mismatches 81; Indels 65; Gaps 8;
Qy	11	LKLAELQOECLARGLETKGIKODLIHLR	-----QAYLEEHAAEEANBEE	54
Db	7	LTVVQLKDLITKRNLSVGGGLKNELVQRLIKDDESKGESEVSPQEQNQOESEPAATIEEP		66
Qy	55	VLGDTEBTEETKPIELPVKGEPEPEKTVDV	-----AAEKVVKITSEIIFQTERMQ	104
Db	67	ASQNIITEKVEVS--SEPKETNEPKENKQVQKPDGSPATASENEQAAAAATAAPALSPEE		124
Qy	105	KRAERFNVPSLESKKAARAARFGISSVPPTKGLSSDNKPMVNLDKLKERAQORFGLNVSS--		163
Db	125	IKAK-----ALDLLNKKLHRANKFG-----	-----QDQADIDSLQQRINRVEKFGVDLNSK	170
Qy	164	-----ISRSKSEDE-----KLKQKRRFQGIVTSGTGTTEDTTEAKRKKAARFG		208
Db	171	LAEBLGHVSRKNPEPSGNGNGKFNRNK-----	NANNRRVSFKNRGRNSG	215

RESULT 9
151116
NF-180 - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 13-Sep-1996 #sequence revision 13-Sep-1996 #text change 09-Jul-2004

C;Accession: I51116
R;Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A;Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A;Reference number: I51116; MUID:95287814; PMID:7770000
A;Accession: I51116
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A;Residues: 1-1110 <JAC>
A;Cross-references: UNIPROT:Q91255; EMBL:U19361; NID:g632548; PIDN:AAA80106.1; PID:g632529
C;Superfamily: neurofilament triplet H protein

Query Match 12.0%; Score 125; DB 2; Length 1110;
Best Local Similarity 24.4%; Pred. No. 2.5;
Matches 46; Conservative 34; Mismatches 85; Indels 26; Gaps 5;

QY 30 IKDLIHRLQAYLEEHAEENBEDVLGDTEBEETKPIELPVKEBEPPEKTVDVAAEKK 89
Db :
660 VESEEAEEAEAEAEAEAEAGEEDVEAESKESEDEDSKDAEDAEEDEAESEEV---KEEE 716

QY 90 VVKITSEIPOTERMQKAERFNVVVSLESK-----KAARAFGISSVPTKGLSS 139
Db :
717 VTKSDAEEAE-AEEEAASKEEAEEKADEAEAEAEAEAEAEAEAEAEAEAEAEAD 775

QY 140 DNKPVMNLDKERAQRFLGNVISISRSDDKLKKRKRFIVTSSAGTGTTDTTEAK 199
Db :
776 DERP-----EEVKESEAPVAPEAKPAEPKPKK-----APAKVESPTSEPEDP 823

QY 200 KRKAERFGIA 210
Db :
824 KAEVWEKKGA 834

RESULT 10
S00030
neurofilament triplet M protein - mouse
N;Alternate names: 160K NF protein; NF-M
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 09-Jul-2004
C;Accession: S00030
R;Levy, E.; Liem, R.K.H.; d'Eustachio, P.; Cowan, N.J.
Eur. J. Biochem. 166, 71-77, 1987
A;Title: Structure and evolutionary origin of the gene encoding mouse NF-M, the middle-m
A;Reference number: S00030; MUID:87246694; PMID:3036526
A;Accession: S00030
A:Molecule type: DNA
A;Residues: 1-849 <LEV>
A;Cross-references: UNIPROT:P08553; GB:X05640; NID:G53357; PIDN:CAA29127.1; PID:g297529
C;Genetics:
A;Map position: 14
A;Introns: 359/3; 401/2
C;Superfamily: cytoskeletal keratin
C;Keywords: coiled coil; intermediate filament
F;99-244/Domain: coiled coil <COI>
F;267-411/Domain: coiled coil <CO2>
F;412-849/Domain: tail <TAI>

Query Match 11.9%; Score 123.5; DB 2; Length 849;
Best Local Similarity 23.6%; Pred. No. 2.3;
Matches 55; Conservative 30; Mismatches 57; Indels 91; Gaps 11;

QY 43 EEHAEEANBEDVLGDTEBEETKPI-----ELPVKEE---EPPEKTVD----- 83
Db EEBGETEAEGE--GEAEAEEKKIEGKVVEAVVAVEEIIVKEKPEAKSPMPKSPVEVK 617

QY 84 -----VAAEKVKVTLSIPTERMQKAERFNVVVSLESKAAAR 126
Db :

Db 618 KPPEAKAGGEQEKEEKVBEEK--EVTKSGPKEEKVEKKEK---PKOVADKKAE--- 669

QY 127 FGISSVPTKGL-----SSDNKPMVNLDLKERAORFG----- 158
Db :

Db 670 ---SPVKEKAVEEVITTSKVKVSLNKDTKKEFPQP-EKVKEKABEEGGSEEGSDRSP 725

	Qy	159	-----LNVSISIRKSDEKLLK---RKRFPGIVTSSAGTGTTEDTEAK	199
	Db	726	QESKKEDIAINGEVGKEEBEQETQEKGSGREEBKGVVITGLDVSPAEEKGE	778
	RESULT 11			
	F87754		protein C43E11.1 [imported] - Caenorhabditis elegans C/Species: Caenorhabditis elegans C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004 C/Accession: F87754 R:anonymous, The C. elegans Sequencing Consortium. Science 282, 2012-2018, 1998 A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biology A/Reference number: A75000; MUID:99069613; PMID:9851916 A/Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/c_elegans/ A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and A/Accession: F87754 A>Status: preliminary A/Molecule type: DNA A/Residues: 1-363 <STO> A/Cross-references: UNIPROT:P91156; GB:chr_I; PIDN:AAB37630.1; PID:g1703576; GSPDB:GN0001 C/Genetics: A/Gene: C43E11.1 A/Map position: 1	
	Query Match	11.8%	Score 123;	DB 2; Length 363;
	Best Local Similarity	25.3%;	Pred. No. 1;	
	Matches	63; Conservative	29; Mismatches	85; Indels 72; Gaps 11;
	Qy	1	MATETV-----ELHLKLAEKQBCLARGLETGKIQLIHRLQ-----AYLEEHAE	47
	Db	1	MADEDILVDGRPLSSLKVTELKELENRLQSLTKGVAVLGQRLEREVILTIKVRFLRVYR	60
	Qy	48	EANEEDVLGDTEETEETKTIELPV-----KEEEP---PEKTVDAAEKKVKVI	93
	Db	61	RRSNQRPQG--REELHGKSIDSVLKFNFDYSNFRKKDOVEAPEAPEDIAPAEPKVAI	118
	Qy	94	TSE-----IPOTERMOKAEREN--VPVSLESKKARAARFGISSVTPTKGLSSDNKPM	144
	Db	119	VKEPKPPEPIETETAAPSRAEKADAAVESVKEKDATAOE---SDSQEAEEKTEPDQ	175
	Qy	145	VNLDKLKERAQRFGLNVSSISRKSEDDEKLKRKER-----FGIVTSSAGTGTTEDT	196
	Db	176	VELS-----NGQBEDDPDRDESIEKQPEKDDGGDELDTYG-----DD	212
	Qy	197	EAKRRKAEE	205
	Db	213	EEKEQKODD	221
	RESULT 12			
	E71169		hypothetical protein PH0553 - Pyrococcus horikoshii C/Species: Pyrococcus horikoshii C/Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004 C/Accession: E71169 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998 A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaelon A/Reference number: A71000; MUID:98344137; PMID:9679194 A/Accession: E71169 A>Status: preliminary; nucleic acid sequence not shown; translation not shown A/Molecule type: DNA A/Residues: 1-399 <RAW> A/Cross-references: UNIPROT:O58288; GB:AF000002; NID:G3236129; PIDN:BAA29642.1; PID:G3256 A/Experimental source: strain OT3 A/Note: this accession replaces an interim accession for a sequence replaced by GenBank C/Genetics: A/Gene: PH0553	
	Query Match	11.8%	Score 122	5; DB 2; Length 399;

Best Local Similarity 26.6%; Pred. No. 1.2;
Matches 54; Conservative 34; Mismatches 68; Indels 47; Gaps 10;
Qy 1 MATETVELHLKLAELKQECIARGL-ETKGIKQDLIHLRLQAYLEEAEE-----48
Db 136 LAKKVEEIHK-KVEELEKLTGKLEETKKVKEELEKIEKGEVTAEEVGELAEKVEEL 194
Qy 49 BAN-----BEDVLGDETEBEETKPI---ELPVKEEPPPEKTVDVAAEKVKVITSEIQTE 101
Db 195 EAAAKPEEEVVVPEEVEEKVEVEEVPPEEVEVTVPEVVEEKEEVEEE-KEKE 253
Qy 102 RMQKRAERFNPVSL-----ESKKAARAARFGISSVPTK-----GLSSDNKPM 144
Db 254 EVGMSEKWKIPEDIAILFEEBPKAR-----LEBIPEDVVSIMIALKWLGLIDRVGI 308
Qy 145 VNLDKLKERAQRFGLNVSSISRK 167
Db 309 QNLEKVLB----FYIEGWISEK 327
RESULT 13
E70474
translation initiation factor IF-2 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ovi
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: E70474
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-805 <AOF>
A:Cross-references: UNIPROT:O67825; GB:AE000769; NID:G2984262; PIDN:AA007794.1; PID:G298
A:Experimental source: strain VF5
C:Genetics:
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu hom
F:309-419/Domain: translation elongation factor Tu homology <ETU>
F:315-322/Region: nucleotide-binding motif A (P-loop)
F:416-419/Region: GTP-binding NKXD motif
F:452-454/Region: GTP-binding SAK/L motif
F:321,322,342,416,417,419,452/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #
Query Match 11.8%; Score 122.5; DB 2; Length 805;
Best Local Similarity 26.3%; Pred. No. 2.5;
Matches 47; Conservative 34; Mismatches 61; Indels 37; Gaps 7;
Qy 22 ARGLETGKIKODLI---HRLQAYLEEAEEANEDVLGDETEE-----ETKPIE 69
Db 56 AFGIKKEEKEEVTEQQAFAEVEEKEEKEEVEEVEEVEEVEEVEEVEEVEEVEE 115
Qy 70 LPVKEEPEPTVDVAAEKVKVITSEIQTERMQKRAERFN-----VPVLSKKA 122
Db 116 EEKKEEKPKEVE-----ELIKETLEKEKEKKEKVEKKEKVEKVEKVEKVEK 170
Qy 123 RAARFGISSVPTKGLSSDNKPMVNLDKLER--AQRFGNLVSSISRKSEDEKLUKKE 179
Db 171 K-----KEEKKEEKPKEKMSK-KEREIMKLEHAKVEKKEKKEKKEKKEKKE 218
RESULT 14
T21174
hypoetical protein F20G4.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T21174; T25113
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19385

A:Accession: T21174
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2020 <WIL>
A:Cross-references: UNIPROT:Q19658; EMBL:Z75538; PIDN:CAA99841.1; GSPDB:GN00019; CESP:F20
A:Experimental source: clone F20G4
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19982
A:Accession: T25113
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2020 <W12>
A:Cross-references: EMBL:Z75550; PIDN:CAA99931.1; GSPDB:GN00019; CESP:F20G4.3
A:Experimental source: clone T22C1
C:Genetics:
A:Gene: CESP:F20G4.3
A:Map position: 1
A:Introns: 133/3; 155/2; 195/2; 270/3; 325/1; 1023/3; 1072/2; 1125/2; 1207/3; 1248/3; 13
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:106-798/Domain: myosin motor domain homology <MMO>
Query Match 11.6%; Score 120.5; DB 2; Length 2020;
Best Local Similarity 24.8%; Pred. No. 8.8;
Matches 56; Conservative 32; Mismatches 91; Indels 47; Gaps 8;
Qy 7 ELHKLKLAEIKQECIARGLETGKIKQDLIHLRLQAYLEEAEEANEDVL-----GDE 59
Db 1153 ETNKEKAARQAE-----KARRDMAELESYKQE--LEESNDKTVLHLSQLKAKRDE 1201
Qy 60 TESEETKPIELPVKEEPPPEKTVDVAAEKVKVITSEIQTERMQK-----105
Db 1202 EYAHQLQLEETVKSSVEEVEEMKAQOKKIBELNETIDQLKR-OKISADKAKSSAESDN 1260
Qy 106 ---RAERFNV-PVSLESKKAARAARFGISSVPTKGLSSDNK---PMVNLDKLKERAQRF 158
Db 1261 ENFRAELSNLSARLAEAEKKAARAE-----TSLMEKDHKWMQSNLDDLMAKLSKMN 1313
Qy 159 LNVSSISRKSEDEDEKLKKRERFGIVTSAGTGTEDTEAKKRKA 204
Db 1314 NELESIQAKSADETLNSLLKKNASLMDQSELTEASEEDRRTRA 1359
RESULT 15
T30330
gelsolin-related protein GRP125 - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30330
R:Stocker, S.; Hiery, M.; Marriott, G.
Mol. Biol. Cell 10, 161-178, 1999
A:Title: Phototactic migration of Dictyostelium cells is linked to a new type of gelsolin
A:Reference number: Z20823; MUID:99096692; PMID:9880334
A:Accession: T30330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1087 <STO>
A:Cross-references: UNIPROT:O96923; EMBL:U95159; NID:G4100185; PID:G4100186; PIDN:AA0007
C:Genetics:
A:Introns: 137/1
Query Match 11.5%; Score 119.5; DB 2; Length 1087;
Best Local Similarity 24.1%; Pred. No. 5.3;
Matches 56; Conservative 38; Mismatches 103; Indels 35; Gaps 9;
Qy 4 ETVELHKLKLAEIKQECIARGLETGKIKQDLIHLRLQAYLEEAEEANEDVLGDETEE 63
Db 724 EKEDLEKLK-QQQEQEQQQKNNKIVVEEVEEK---EDVKKEEVEEVEEVEEKEE 779
Qy 64 ETKPIELPVKEEPPPEKTVDVAAE-KKVKVITSEIQTE-----RMQKRAERFNPVPSLE 117
Db 780 EVKEVAKEETKEEIKKEEVNDEATEVKEVQVVEEVEEVEEVEEVEEVEEVEEVEE 839

```
Qy 118 SKAARAARFGISSVPTKGLSSDNKPMVNLDKLKERAZQRFGLNVSSISRKSED--DEKL 174
Db 840 EVKEEVKEEVKE-EVKEVKEEVKEEVKQKKEEVNEEIKKEETKEETKEDDNKEDKV 898
Qy 175 KKRKE-----RFGIVTS-----SAGTGTITTEAK-----KKRAE 205
Db 899 NEENETVNEENEVGIIVSPPEKVDSEANSSSTISSPENEGSVSVKDKRKSNE 950
```

Search completed: October 5, 2005, 05:59:36
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 00:10:27 ; Search time 87 Seconds
(without alignments)
1236.053 Million cell updates/sec

Title: US-09-788-476A-2
Perfect score: 1040
Sequence: 1 MATETVELHKLALBKQEC.....GTTEDTEAKKRAERFGIA 210

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1035	99.5	209	1 HCC1 HUMAN	P82979 homo sapien
2	990	95.2	209	1 HCC1 MOUSE	Q9dlj3 mus musculus
3	330	31.7	756	2 Q6Y4V2	Q6Y4V2 bovine vira
4	276.5	26.6	363	2 Q7QBC8	Q7QBC8 anopheles g
5	222	21.3	245	1 YCM3 SCHPO	O74871 schizosacch
6	222	21.3	253	2 Q9N3G0	Q9N3G0 caenorhabdi
7	222	21.3	308	2 Q9VHC8	Q9VHC8 drosophila
8	165.5	15.9	209	2 Q967H3	Q967H3 hydra atten
9	155.5	15.0	246	2 Q6C2C7	Q6C2C7 yarrowia li
10	155	14.9	1271	2 Q9V8K6	Q9V8K6 drosophila
11	151.5	14.6	203	2 Q6DEQ1	Q6DEQ1 xenopus tro
12	148	14.2	344	2 Q9P6W5	Q9P6W5 neurospora
13	144	13.8	784	2 Q6NUT5	Q6NUT5 brachydanio
14	143	13.8	177	2 Q8H8U3	Q8H8U3 oryza sativ
15	139	13.4	214	2 Q9LZ08	Q9LZ08 arabidopsis
16	139	13.4	305	2 Q8I327	Q8I327 plasmodium
17	139	13.4	1222	2 Q9XTH8	Q9XTH8 caenorhabdi
18	137.5	13.2	204	2 Q8BIQ7	Q8BIQ7 mus musculus
19	137.5	13.2	1150	2 Q6X935	Q6X935 homo sapien
20	136.5	13.1	230	2 Q6FPLO	Q6FPLO candida gla
21	136.5	13.1	718	2 Q9NV00	Q9NV00 homo sapien
22	136.5	13.1	1150	2 Q8IX12	Q8IX12 homo sapien
23	135.5	13.0	329	2 Q6GQC9	Q6GQC9 xenopus lae
24	135.5	13.0	18519	2 Q8ISR6	Q8ISR6 caenorhabdi
25	135.5	13.0	18534	2 Q8ISE7	Q8ISE7 caenorhabdi
26	135	13.0	774	2 Q9VHY5	Q9VHY5 xenopus lae
27	134.5	12.9	10578	2 Q8ISF5	Q8ISF5 caenorhabdi
28	133.5	12.8	315	2 Q8GL99	Q8GL99 streptococc
29	133.5	12.8	317	2 Q8GL96	Q8GL96 streptococc
30	133.5	12.8	633	2 Q6S655	Q6S655 arabidopsis
31	133	12.8	843	2 Q6AXC9	Q6AXC9 mus musculus

32	133	12.8	1146	2	Q8CH18	Q8CH18 mus musculu
33	131	12.6	719	2	Q7S8Q9	Q7S8Q9 neurospora
34	130.5	12.5	314	2	Q8GLB0	Q8GLB0 streptococc
35	130.5	12.5	402	2	Q8U3W6	Q8U3W6 pyrococcus
36	129	12.4	539	2	Q8BJY0	Q8BJY0 m mus muscu
37	128	12.3	1157	2	Q641G3	Q641G3 xenopus lae
38	128	12.3	2006	2	Q7KSQ6	Q7KSQ6 plasmodium
39	128	12.3	2019	2	Q7KSQ5	Q7KSQ5 plasmodium
40	128	12.3	2055	2	Q8IHP3	Q8IHP3 plasmodium
41	128	12.3	2055	2	Q8TSC7	Q8TSC7 plasmodium
42	127.5	12.3	804	2	Q6PIZ0	Q6PIZ0 homo sapien
43	126.5	12.2	326	2	Q8GL91	Q8GL91 streptococc
44	126.5	12.2	580	2	P91156	P91156 caenorhabdi
45	126.5	12.2	809	2	Q6PAR2	Q6PAR2 mus musculu

ALIGNMENTS

RESULT 1

ID HCC1 HUMAN STANDARD; PRT; 209 AA.

AC P82979; Q9P066;

DT 16-OCT-2001 (Rel. 40, Created)

DT 25-OCT-2004 (Rel. 45, Last sequence update)

DE Nuclear protein Hcc-1 (HSPC316) (Proliferation associated cytokine-inducible protein CIP29).

GN Name=HCC1;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.; SEQUENCE OF 109-118; 156-166 AND 180-198, TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.

RC TISSUE=Liver;

RX MEDLINE=21255661; PubMed=11356193; DOI=10.1016/S0014-5793(01)02409-7; Choong M.L., Tan L.K., Lo S.L., Ren E.-C., Ou K.L., Ong S.-E., Liang R.C.M.Y., Seow T.K., Chung M.C.M.;

RA "An integrated approach in the discovery and characterization of a novel nuclear-protein-over-expressed in liver and pancreatic tumors."

RL FBS Lett. 496:109-116(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Umbilical cord blood;

RX MEDLINE=20499367; PubMed=11042152; DOI=10.1101/gr.140200; Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G., Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W., Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;

RA "Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells."

RL Genome Res. 10:1546-1560(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast cancer;

RX MEDLINE=21920347; PubMed=11922608; DOI=10.1006/bbrc.2002.6680; Fukuda S., Wu D.W., Stark K., Pelus L.M.;

RA "Cloning and characterization of a proliferation-associated cytokine-inducible protein, CIP29."

RL Biochem. Biophys. Res. Commun. 292:593-600(2002).

RN [4]

RP SEQUENCE FROM N.A.

RC TISSUE=Prostate;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899; Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Scapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RN SEQUENCE OF 1-9 AND 126-134, AND ACETYLATION SITE ALA-1.
RN Blenvenut W.V., Potts A., Brablan J., Quadroni M.;
RA Unpublished observations (JUI-2004).
CC -1- FUNCTION: May have nucleic acid binding capability that may
CC participate in important transcriptional or translational control
CC of cell growth, metabolism and carcinogenesis.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Low expression in spleen, liver, pancreas,
CC testis, thymus, heart, and kidney. Increased levels are seen in
CC hepatocellular carcinoma and pancreatic adenocarcinoma.
CC -1- INDUCTION: By erythropoietin.
CC -1- SIMILARITY: Contains 1 SAP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; A409089; CAC37950.1; -
DR ENBL; AF161434; AAF28994.1; ALT_INIT.
DR ENBL; AF486281; AAM09686.1; -
DR ENBL; BC007099; AAH07099.1; -
DR PIR; JC7830; JC7830.
DR IntAct; P82979; -
DR H-InvDB; HIX0010706; -
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003676; F:nucleic acid binding; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR GO; GO:0006445; P:regulation of translation; NAS.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR PROSITE; PS50800; SAP; 1.
KW Acetylation; Direct protein sequencing; DNA-binding; Nuclear protein;
KW Phosphorylation; Transcription; Transcription regulation;
KW Translation regulation.
FT INIT MET 0 0
FT MOD_RES 1 1 N-acetylalanine.
FT DOMAIN 7 41 SAP.
FT MOD_RES 162 162 Phosphoserine (By similarity).
FT CONFLICT 126 126 F -> V (in Ref. 2).
FT CONFLICT 133 148 TGLSSDNKEWNLDK -> NKRSVILITNLWLTWIS (in
FT Ref. 2).
FT CONFLICT 152 156 RAQRF -> ELKDL (in Ref. 2).
FT CONFLICT 198 209 KKKRAERFGIA -> RRGKEQALGPDPEKFLILSVLQCF
FT PFLSFFLATVMPKCTVMCLRPASQ (in Ref. 2).
FT SEQUENCE 209 AA; 23539 MW; F24C5B1CF67C073A CRC64;
Query Match 99.5%; Score 1035; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. NO. 3.1e-51;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ATETVELHKLKLAELKQELCLARGLGTGKIQDILHRLQAVLEEHABEEANEEDVLGDTE 61
Db 1 ATETVELHKLKLAELKQELCLARGLGTGKIQDILHRLQAVLEEHABEEANEEDVLGDTE 60
QY 62 EETKPIELPVKEEPEPTVDVAAEKVKVITSEIPQTERMQKRAERFNVPSLESKKA 121

Db 61 EETKPIELPVKEEPEPTVDVAAEKVKVITSEIPQTERMQKRAERFNVPSLESKKA 120
QY 122 ARAARFGISVPTKGLSSDNKPNVNLDKLKERQRFGLNVSSIRKSXEDDEKLUKKRKERF 181
Db 121 ARAARFGISVPTKGLSSDNKPNVNLDKLKERQRFGLNVSSIRKSXEDDEKLUKKRKERF 180
QY 182 GIVTSSAGTGTTEDEAKKRAERFGIA 210
Db 181 GIVTSSAGTGTTEDEAKKRAERFGIA 209
RESULT 2
HCC1 MOUSE STANDARD; PRT; 209 AA.
ID HCC1_MOUSE STANDARD; PRT; 209 AA.
AC Q9DIJ3; Q9CU18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Nuclear protein Hcc-1.
GN NamesHcc1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, and NOD; TISSUE=Cerebellum, Embryo, and Thymus;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RX Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nigaki I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schraml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gusticich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Ienhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sanderlin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RT Nature 420:563-573(2002).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Mammary gland;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RN SEQUENCE OF 1-9 AND 126-134, AND ACETYLATION SITE ALA-1.
RN Blenvenut W.V., Potts A., Brablan J., Quadroni M.;
RA Unpublished observations (JUI-2004).
CC -1- FUNCTION: May have nucleic acid binding capability that may
CC participate in important transcriptional or translational control
CC of cell growth, metabolism and carcinogenesis.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Low expression in spleen, liver, pancreas,
CC testis, thymus, heart, and kidney. Increased levels are seen in
CC hepatocellular carcinoma and pancreatic adenocarcinoma.
CC -1- INDUCTION: By erythropoietin.
CC -1- SIMILARITY: Contains 1 SAP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; A409089; CAC37950.1; -
DR ENBL; AF161434; AAF28994.1; ALT_INIT.
DR ENBL; AF486281; AAM09686.1; -
DR ENBL; BC007099; AAH07099.1; -
DR PIR; JC7830; JC7830.
DR IntAct; P82979; -
DR H-InvDB; HIX0010706; -
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003676; F:nucleic acid binding; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR GO; GO:0006445; P:regulation of translation; NAS.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR PROSITE; PS50800; SAP; 1.
KW Acetylation; Direct protein sequencing; DNA-binding; Nuclear protein;
KW Phosphorylation; Transcription; Transcription regulation;
KW Translation regulation.
FT INIT MET 0 0
FT MOD_RES 1 1 N-acetylalanine.
FT DOMAIN 7 41 SAP.
FT MOD_RES 162 162 Phosphoserine (By similarity).
FT CONFLICT 126 126 F -> V (in Ref. 2).
FT CONFLICT 133 148 TGLSSDNKEWNLDK -> NKRSVILITNLWLTWIS (in
FT Ref. 2).
FT CONFLICT 152 156 RAQRF -> ELKDL (in Ref. 2).
FT CONFLICT 198 209 KKKRAERFGIA -> RRGKEQALGPDPEKFLILSVLQCF
FT PFLSFFLATVMPKCTVMCLRPASQ (in Ref. 2).
FT SEQUENCE 209 AA; 23539 MW; F24C5B1CF67C073A CRC64;
Query Match 99.5%; Score 1035; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. NO. 3.1e-51;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ATETVELHKLKLAELKQELCLARGLGTGKIQDILHRLQAVLEEHABEEANEEDVLGDTE 61
Db 1 ATETVELHKLKLAELKQELCLARGLGTGKIQDILHRLQAVLEEHABEEANEEDVLGDTE 60
QY 62 EETKPIELPVKEEPEPTVDVAAEKVKVITSEIPQTERMQKRAERFNVPSLESKKA 121

RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmitz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP PHOSPHORYLATION SITE SER-162.
RX PubMed:14729942; DOI=10.1074/mcp.D300003-MCP200;
RA Shu H., Chen S., Bi Q., Mumby M., Brekken D.L.;
RT "Identification of phosphoproteins and their phosphorylation sites in
RT the WHI-231 B lymphoma cell line";
RL Mol. Cell. Proteomics 3:279-286(2004).
CC -!- FUNCTION: May have nucleic acid binding capability that may
CC participate in important transcriptional or translational control
CC of cell growth, metabolism and carcinogenesis (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 SAP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK003453; BAB22799.1; -
DR EMBL; AK018773; BAB31400.1; -
DR EMBL; AK088266; BAC40246.1; -
DR EMBL; BC027510; AAH27510.1; -
DR MGD; MGI:1913368; 1110005A23rik.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
KW Acetylation; DNA-binding; Nuclear protein; Phosphorylation;
KW Transcription; Transcription regulation; Translation regulation.
FT INIT MET 0 0 By similarity.
FT MOD RES 1 1 N-acetylalanine (By similarity).
FT DOMAIN 7 41 SAP.
FT MOD RES 162 162 Phosphoserine.
FT MOD RES 209 AA; 23401 MW; 719300BAE5C76B7 CRC64;
SQ SEQUENCE 209 AA; 23401 MW; 719300BAE5C76B7 CRC64;
Query Match 95.2%; Score 990; DB 1; Length 209;
Best Local Similarity 95.7%; Pred. No. 1.1e-46;
Matches 200; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Qy 2 ATETVELHKLKLAELKQELCLARGLETGKIQDLIHLRQAYLEHAEENEEDVLGDETE 61
Db 1 AAETVELHKLKLAELKQELCLARGLETGKIQDLIHLRQAYLEHAEENEEDVLGDETE 60
Qy 62 EETKPTIELPKKEEPEKTVDAEAKVKVITSEIPQTERMQRAERFNPVVSLESKKA 121
Db 61 EEPKPTIELPKKEEPEKTVDAEAKVKVITSEIPQTERMQRAERFNPVVSLESKKA 120
Qy 122 ARAARFGISSVPTKGLSSDNKPMVNLKDKERAQRFGLNVSSISRKSEDDKLRKRF 181
Db 121 ARAARFGISSVPTKGLSSDNKPMVNLKDKERAQRFGLNVSSISRKSEDDKLRKRF 180
Qy 182 GIVTSSAGTGTTEDTEAKRRAERFGIA 210
Db 181 GIVTSSAGTGTTEDTEAKRRAERFGIA 209
RESULT 3
Q6Y4V2 PRELIMINARY; PRT; 756 AA.
AC Q6Y4V2
DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Polyprotein (Fragment).
OS Bovine viral diarrhea virus (BVDV) (Mucosal disease virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_TaxID=11099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CP8;
RX PubMed:12970452; DOI=10.1128/JVI.77.19.10663-10669.2003;
RA Mueller A., Rinck G., Thiel H.-J., Tautz N.;
RT "Cell-derived sequences in the N-terminal region of the polyprotein of
RT a cytopathogenic pestivirus";
RL J. Virol. 77:10663-10669(2003).
DR EMBL; AY182136; AAC86641.1; -
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR GO; GO:0019082; P:viral protein processing; IEA.
DR InterPro; IPR008751; Peptidase_C53.
DR InterPro; IPR003034; SAP.
DR Pfam; PF05550; Peptidase_C53; 1.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
KW Polyprotein.
FT NON TER 756 756
SQ SEQUENCE 756 AA; 85191 MW; 71D0EE9E05B669D8 CRC64;
Query Match 31.7%; Score 330; DB 2; Length 756;
Best Local Similarity 37.2%; Pred. No. 7.7e-11;
Matches 97; Conservative 24; Mismatches 68; Indels 72; Gaps 8;
Qy 12 KLAELKQELCLARGLETGKIQDLIHLRQAYLEHAEENEEDVLGDETEETKPTIELP 71
Db 445 KLAELKQELCLARGLETGKIQDLIHLRQAYLEHAEENEEDVLGDETEETKPTIELP 504
Qy 72 VKEEPEPEKTVDAEAKVKVITSEIPQTE 101
Db 505 VKESARQGTAKLTWLGRLGILKKLEEPVVDQAGNPLFGRGVHPQSTLKLPL 561
Qy 102 RMQKRAERFNPVVSLESKKAARAFGISVPTKGLSSDNKPMVNL 150
Db 562 --HKRGER-EVPTNLASLPKRGDCRSNDKGPVSGIYLPKGPLFYQDYEGPVYHRAPLEL 618
Qy 151 KEPA-----QRGLNVSSISR-----KSEDEKLKK---RKRFQIVT 185
Db 619 FEETATWCEVTKRIGRVGTGSDSKLYHYVICDGCIIKVSATKDKRQVLRKWNKNCPLWV 678
Qy 186 SSAGTGTTEDTEAKRRAER 206
Db 679 SSCSDTKDEGAVRKKQKQKPD 699
RESULT 4
Q7QBC8 PRELIMINARY; PRT; 363 AA.
AC Q7QBC8
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AGCP2865 (Fragment).
GN Name=agCS5572; ORFNames=ENSANGG00000012227;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an

RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537558;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Swirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Swirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Battencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003798; AAF57660.2; -.
DR IntAct; Q9V8K6; -.
DR FlyBase; FBgn0050122; CG30122.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003034; SAP.
DR InterPro; IPR003877; SPRY_receptor.
DR Pfam; PF02037; SAP; 1.
DR Pfam; PF00622; SPRY; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
SQ SEQUENCE 1271 AA; 14049 MW; 322913D560786202 CRC64;

Query Match 14.9%; Score 155; DB 2; Length 1271;
Best Local Similarity 21.08; Pred. No. 1;
Matches 55; Conservative 49; Mismatches 86; Indels 72; Gaps 8;

QY 4 ETVELHKLAEKQECIARGLETGKIQDLHRLQAYLE-----EHA----- 46
DB 2 DVAKLEKMKVVDLRLNELQSGLDTKGVKAVLVERLRAVYEGGAGDGENAPVTSRQRRT 61
QY 47 -----EEANEDVLGDETEETEETKPIELPVKEPEPEKTV 82
DB 62 RSMRSRSPVQAAFPVAAPVLDLTLEEEQEDKTVPOPESESQPAAEPEQEPEEAE 121
QY 83 DVAAEKVKVITSEI-----PQTERMQKRAERFNVPSLESKKAARAGIGSSVPTK-- 135
DB 122 PAAAVTDTTQNAQVNEESQPEPEFDEKSETDDKQETIEAVPVAVVQNEVADEPWEED 181
QY 136 --GLSSDNKPMVNLDKLRAQRFGLNVSSIS--RKSEDDKLGKRGKRFGIVTSAGTG 191

DB 182 HDAAPBEQPTQTETPEVEEKPAE-----STVAHQSGDQKMDVDE-----DSAAPK 230
QY 192 TTIEDTE-----AKKKRA 204
DB 231 TAETETPAAKPEDQPPERRKRS 252

RESULT 11
Q6DEQ1 PRELIMINARY; PRT; 203 AA.
ID Q6DEQ1
AC Q6DEQ1; 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE HNRPU protein (Fragment).
GN Names=HNRPU;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC077044; AAH77044.1; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
FT NON_TER 203
SQ SEQUENCE 203 AA; 21594 MW; 76842B4E18B0CF0E CRC64;

Query Match 14.6%; Score 151.5; DB 2; Length 203;
Best Local Similarity 28.7%; Pred. No. 0.23;
Matches 45; Conservative 25; Mismatches 40; Indels 47; Gaps 5;

QY 6 VELHUKLAEKQECIARGLETGKIQDLHRLQAYLEHAE-----EEAN----- 51
DB 5 INVKKLVSELKEELKRNLSKGLKADLMERLQAALDEGSGAGGGVAVPDEADLGD 64
QY 52 -EDVLGDETEETEETKPIELPVK-----EEPPKTVDAAEKVKVITSETPQTER-- 102
DB 65 PQGATGDEEEEBEGMELGGENGDDGADEAPDEVP-----GEEVPEAEQGD 114
QY 103 -----MOKRAERFNVPSLESKKAARAA 125


```
Db 115 ENGDDQGFQEGDEEEDEEGIPVGLSEBDDAEAA 151
RESULT 12
Q9P6W5 PRELIMINARY; PRT; 344 AA.
AC Q9P6W5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein 68B2.110.
GN Name=68B2.110;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL353821; CAB88619.1; -.
DR PIR; T48827; T48827.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 344 AA; 36133 MW; 5510625B8CAF16343 CRC64;
Query Match 14.28; Score 148; DB 2; Length 344;
Best Local Similarity 27.28; Pred. No. 0.63; Indels 44; Gaps 8;
Matches 62; Conservative 27; Mismatches 95;
QY 7 ELHKLAELKQECCLARGLETGKIGKQDLIHLQAYLEHAEHEEANEEDVL-----GDE 59
DB 3 DYNMKVPKELKLLNLSLPTGNKADLIARLQ-----EHDKQAKPADAPATTKDGEA 58
QY 60 TEEETKPIELPVKEEPPPEKTVDAAEKVKVITSEIPOTERMOKRAERNVPVLSK 119
DB 59 EDEIDYDDDFPAGDKKAATETD---EKAPASAAAATATSESPK-----ABVESEAQ 109
QY 120 KAARAAAFGISSVPTKGLSDNKPVMNL-----DKLKERAR-----FGLNVS 162
DB 110 KTTAAAPAAAASTQPAASQAQGEVAAKOTTTATATTAEEDPSKAEQKPAEPLFSQHLP 169
QY 163 SISRKSDDEKLKKRERFGVITSSAGTGTTEDEAKKRAERFGIA 210
DB 170 PTDKSE-AKRAARAARFGITDE-----KSEEAQAKAARAARFGIA 210
RESULT 13
Q6NUT5 PRELIMINARY; PRT; 784 AA.
AC Q6NUT5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Zgc:85971.
GN ORNames=zgc:85971;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
```

```
RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Buell R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084405; AAN64490.1; -
DR Gramene; Q8H8U3; -
KW Hypothetical protein.
SQ SEQUENCE 177 AA; 18264 MW; B90FF1D5D0500C8A CRC64;

Query Match      13.8%; Score 143; DB 2; Length 177;
Best Local Similarity 39.4%; Pred. No. 0.6;
Matches 43; Conservative 14; Mismatches 32; Indels 20; Gaps 5;

QY 101 ERMQKRAEREN--VPVSLSKKAARAFGISSVPTKGLSSDNKPMVNLDKLKERAQRFG 158
   | : ||||| | : | : ||||| | : | : ||||| | : | : ||||| | : | : |||||
Db 58 EKQMRRAERPGTAVVSEEEKRSRAERFGTG-----SSNEK--AEQKRSRAERFG 108
   | : ||||| | : | : ||||| | : | : ||||| | : | : ||||| | : | : |||||

QY 159 LVNSSISRSKSEDEKLKKRKRFIVTSSAGTGTGTTEDTEAKRKRAERF 207
   | : ||||| | : | : ||||| | : | : ||||| | : | : ||||| | : | : |||||
Db 109 LASSS----ADEDAKKARLERFG-----QSTNVDRGEEKRKARALRF 148

RESULT 15
Q9LZ08 PRELIMINARY; PRT; 214 AA.
ID Q9LZ08
AC Q9LZ08;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein F9G14_80 (AT5G02770/F9G14_80).
GN Names=F9G14_80;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terryn N., Ardiles W., Buyshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van den Daele H.,
RA Villarroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Chung M.K., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yu G., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162973; CAB86031.1; -

DR EMBL; AY050475; AAK91488.1; -
DR EMBL; AF378893; AAKS5696.1; -
DR FIR; T48298; T48298.
KW Hypothetical protein.
SQ SEQUENCE 214 AA; 22779 MW; 2307B981E9297ADE CRC64;

Query Match      13.4%; Score 139; DB 2; Length 214;
Best Local Similarity 31.2%; Pred. No. 1.2;
Matches 55; Conservative 22; Mismatches 67; Indels 32; Gaps 7;

QY 57 GDETEEETKPIELPVKEEPEPEKTVD-----VAAEKVVVKITSEIPOT----- 100
   | : ||||| | : | : ||||| | : | : ||||| | : | : ||||| | : | : |||||
Db 17 GLSTGENPKIIVDLNTTDLTDILDGKVGKGFSDSGEKKETDSNGIGSTAGVDSGDIS 76

QY 101 -----ERMQKRAERENVPVSL--ESKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKER 153
   | : ||||| | : | : ||||| | : | : ||||| | : | : ||||| | : | : |||||
Db 77 PVDDIQKKIRRAERFGSVKLTTEEKRNRAERFGTVAADVNGSEGTGKKAEL--KRKA 135

QY 154 AORFGI--NVSSISRSKSEDEKLKKRKRFIVTSSAGTGTGTTEDTEAKRK--RAERF 207
   | : ||||| | : | : ||||| | : | : ||||| | : | : ||||| | : | : |||||
Db 136 ADRFGVPSATSTTDKTEEAKKKARLAFGKTK-----VDSAEENKKARALRF 185

Search completed: October 5, 2005, 05:59:03
Job time : 97 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

• OM protein - protein search, using sw model

Run on: October 5, 2005, 05:33:55 ; Search time 29 Seconds
(without alignments)
540.562 Million cell updates/sec

Title: US-09-788-476A-2
Perfect score: 1040
Sequence: 1 MATETVELHKLKLAELKQEC.....GTTEDTEAKRKRAEREGIA 210

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	742	71.3	149	4	US-09-513-999C-4813
2	625	60.1	139	4	US-09-621-976-7487
3	143.5	13.8	343	4	US-09-270-767-42868
4	125.5	12.1	218	4	US-09-538-092-241
5	120.5	11.6	824	4	US-09-538-092-1242
6	120.5	11.6	1180	4	US-09-543-681A-6436
7	116	11.2	806	1	US-07-980-528-2
8	115	11.1	428	4	US-09-286-981B-7
9	114.5	11.0	253	4	US-09-270-767-41091
10	114.5	11.0	253	4	US-09-270-767-56307
11	114.5	11.0	1507	3	US-08-923-329-5
12	113.5	10.9	1089	4	US-09-949-016-10326
13	113	10.9	413	4	US-09-286-981B-5
14	112	10.8	817	4	US-09-248-796A-20276
15	111	10.7	406	4	US-09-286-981B-18
16	111	10.7	667	4	US-09-248-796A-14492
17	110.5	10.6	375	4	US-09-538-092-1082
18	110.5	10.6	375	4	US-09-949-016-6369
19	110.5	10.6	386	4	US-09-949-016-10070
20	110.5	10.6	1402	4	US-09-248-796A-14503
21	109	10.5	735	4	US-10-164-595-80
22	109	10.5	779	4	US-10-164-595-56
23	109	10.5	784	4	US-10-164-595-79
24	109	10.5	843	4	US-10-164-595-54
25	108.5	10.4	1231	4	US-08-714-741-41
26	108	10.4	588	4	US-08-714-741-42
27	108	10.4	864	4	US-08-714-741-40

28	107.5	10.3	447	4	US-09-710-279-2900	Sequence 2900, Ap
29	107.5	10.3	450	4	US-09-710-279-3226	Sequence 3226, Ap
30	107.5	10.3	2079	4	US-09-949-016-8301	Sequence 8301, Ap
31	107	10.3	412	2	US-08-741-134-2	Sequence 2, Appli
32	107	10.3	440	4	US-09-762-194-2	Sequence 2, Appli
33	107	10.3	440	4	US-09-762-194-4	Sequence 4, Appli
34	106	10.2	477	4	US-09-248-796A-15439	Sequence 15439, A
35	106	10.2	529	4	US-09-248-796A-16357	Sequence 16357, A
36	106	10.2	1702	3	US-08-296-791-5	Sequence 5, Appli
37	106	10.2	1702	4	US-09-839-996-5	Sequence 5, Appli
38	106	10.2	1702	4	US-10-080-505-5	Sequence 5, Appli
39	106	10.2	1702	4	US-10-645-655-5	Sequence 5, Appli
40	106	10.2	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
41	106	10.2	1786	3	US-08-973-462-8	Sequence 8, Appli
42	105.5	10.1	1076	4	US-09-976-594-889	Sequence 889, App
43	105.5	10.1	1857	4	US-09-917-254-91	Sequence 91, Appl
44	105.5	10.1	1972	4	US-09-538-092-1084	Sequence 1084, Ap
45	105.5	10.1	1984	4	US-09-949-016-7111	Sequence 7111, Ap

ALIGNMENTS

RESULT 1
US-09-513-999C-4813
; Sequence 4813, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US/60/122,487
; NUMBER OF SEQ ID NOS: 3668
; SOFTWARE: Patent.pm
; SEQ ID NO 4813
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-4813

Query Match	71.3%	Score 742;	DB 4;	Length 149;
Best Local Similarity	100.0%	Pred. No. 5.2e-62;		
Matches 149;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MATETVELHKLKLAELKQEC	1	MATETVELHKLKLAELKQEC
Db	1	MATETVELHKLKLAELKQEC	1	MATETVELHKLKLAELKQEC
Qy	61	EEETKTIELPVKEEPEPEKTDVAAEKVKVITSEIPQTERMOKRAERFNPVVSLESKK	120	EEETKTIELPVKEEPEPEKTDVAAEKVKVITSEIPQTERMOKRAERFNPVVSLESKK
Db	61	EEETKTIELPVKEEPEPEKTDVAAEKVKVITSEIPQTERMOKRAERFNPVVSLESKK	120	EEETKTIELPVKEEPEPEKTDVAAEKVKVITSEIPQTERMOKRAERFNPVVSLESKK
Qy	121	AARAARFGISVPTKGLSSDNKPMVNLDK	149	AARAARFGISVPTKGLSSDNKPMVNLDK
Db	121	AARAARFGISVPTKGLSSDNKPMVNLDK	149	AARAARFGISVPTKGLSSDNKPMVNLDK

RESULT 2
US-09-621-976-7487
; Sequence 7487, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 7487
; LENGTH: 139
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-621-976-7487

Query Match 60.1%; Score 625; DB 4; Length 139;
Best Local Similarity 100.0%; Pred. No. 4.2e-51;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATETVELHLKLAELKQECCLARGLETGKIGIKODLIHRLQAYLEHAEHEEANEEDVLGDET 60
DB 1 MATETVELHLKLAELKQECCLARGLETGKIGIKODLIHRLQAYLEHAEHEEANEEDVLGDET 60
QY 61 EEETKPIELPVKEEPEPEKTDVAEEKVVKITSEIPQTERMQKRAERNVPSLESKK 120
DB 61 EEETKPIELPVKEEPEPEKTDVAEEKVVKITSEIPQTERMQKRAERNVPSLESKK 120
QY 121 ARAAR 126
DB 121 ARAAR 126

RESULT 3
US-09-270-767-42868
; Sequence 42868, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42868
; LENGTH: 343
; TYPE: PR1
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-42868

Query Match 13.8%; Score 143.5; DB 4; Length 343;
Best Local Similarity 24.4%; Pred. No. 1.7e-05;
Matches 61; Conservative 39; Mismatches 81; Indels 69; Gaps 10;
QY 7 ELHLKLAELKQECCLARGLETGKIGIKODLIHRLQA-----YLEEHAEEANEED----- 54
DB 18 DVTMKVADLKRELKRLGLAVNGNKTQLDRQLTALLEGDLSDLSAIDADDVVSFT 77
QY 55 -----VLGDETEEEETK-PIELPVKEEPEPEKTDVAEEKVVKITSEIPQTERMQK- 106
DB 78 DEDEHLLGDNDDELKPSVSTPTTVAIP-----DLAEEK-----TSSAPDAAAPTCKI 128
QY 107 -----AERFNVPSLESKKAARAFGIVSSVPTKGLSSDNKPMV----- 145
DB 129 VLKRNNSQOSTGTVASTGTPP-SKENEAPAAASDSTGETPTK-----KHKPIVVGPKTEG 183
QY 146 -----NLDKLKERAQFGLNVSSISRKSDDEDEKLKKRKRFIVTSSAGTGT 192
DB 184 EKPSGDKKLNLTQAERLELRKAKFGITTPPAVA-NTATAVAVAINKSSASITANKNGKE 242
QY 193 TEDTEAKKRK 202
DB 243 TEOQKEASK 252

RESULT 4

US-09-538-092-241
; Sequence 241, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormat Version 0.9
; SEQ ID NO 241
; LENGTH: 218
; TYPE: PR1
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YER063W
US-09-538-092-241

Query Match 12.1%; Score 125.5; DB 4; Length 218;
Best Local Similarity 22.0%; Pred. No. 0.00044;
Matches 52; Conservative 38; Mismatches 81; Indels 65; Gaps 8;
QY 11 LKLAELKQECCLARGLETGKIGIKODLIHRL-----QAYLEHAEHEEANEED 54
DB 7 LTVVQKLDLTKNLVSGGLKNELVORLIKDDSEKGESEVSPQEQNOEGSPAAIEEP 66
QY 55 VLGDETEEBETKPIELPVKEEPEPEKTDV-----AAKKVVKITSEIPQTERMQ 104
DB 67 ASQNIETKEKVS--SEPKETNEPKENKDVQKPSDGPSTASENEQAAASTAAPALSPEE 124
QY 105 KRAERNVPSLESKKAARAFGIVSSVPTKGLSSDNKPMVNLKLERAQRTGLNVSS- 163
DB 125 IKAK-----ALDLINKKLHRAKFG-----QDQADIDSQRIQINRVEKFGVDLNSK 170
QY 164 -----ISKSDDE-----KLKKRKRFIVTSSAGTGTTEAKKRKRAERFG 208
DB 171 LAELGLVSRKNEPESGNGKFKRNK-----NANNRSRVSKRRGRNSG 215

RESULT 5
US-09-538-092-1242
; Sequence 1242, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormat Version 0.9
; SEQ ID NO 1242
; LENGTH: 824
; TYPE: PR1
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q00839
US-09-538-092-1242

```

; APPLICANT: Kiledjian, Megerditch
; APPLICANT: Portman, Douglas S.
; TITLE OF INVENTION: METHODS OF PROMOTING INTERMOLECULAR
; INTERACTIONS INVOLVING A NUCLEIC ACID AND MOLECULES
; TITLE OF INVENTION: USEFUL
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5457026ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25; mb/MD
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/980,528
; FILING DATE: 199211120
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Debuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPN-0847
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 806 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-980-528-2

Query Match 11.2%; Score 116; DB 1; Length 806;
Best Local Similarity 23.4%; Pred. No. 0.019;
Matches 64; Conservative 37; Mismatches 82; Indels 90; Gaps 10;

Qy 1 MATETVELHKLKLAELKQECCLARGLTGTGKQDLIHLQAYLEH-----ABEEAN--- 51
Db 1 MSSSPVNVKLVKSELKELKRLSDGLKALMERLQAALDDEEAGGRPAMEPGNSL 60
Qy 52 -----EEDVLGDTEETETK-----PIEL-----P 71
Db 61 DLGDSAGRSAGLGLEQEAAGGDEEEEEESEEGISALDQOMELGEENGAAGAASGP 120
Qy 72 VKEEE---PPEKTVDAAEKKVVKITSE-----IPOTERMQKRAERFNV 112
Db 121 MESEEAASEDNGDDQGFQGEDELDEEERGAGDENGHGQEQPPATQOQPPQQRG-- 178
Qy 113 PVSLESKKAARAARFGISSYPTKGLSDNDKPMVNLDKLKERARFGLNVSSISRSBDE 172
Db 179 ----AAKEAA-----GKSSGPTSLFAVTVP-----PGARQQQQAAGDGKTEQKGGDKR 225
Qy 173 KLKRRKERFIVTSSAGTGTTEDTEAKKRRAE 205
Db 226 GVKRRPRE-----DHGRGYFEYIEENKYSRAK 251

RESULT 8
US-09-286-981B-7
; Sequence 7, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

```

```
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-7

Query Match      11.1%; Score 115; DB 4; Length 428;
Best Local Similarity 26.0%; Pred. No. 0.01;
Matches 59; Conservative 28; Mismatches 90; Indels 50; Gaps 10;

QY      2 ATETVELHKLKLAELKQEC--LARGLETGKIQDLIHLRLQAYLE-----EHAEEANEED 54
Db      165 ANEPRDEKIQAEKVESQAEATRLKKIKTDREQAETRLNIKTDRQAEEBAKVKD 224
QY      55 -----VLGD-----ETBEETKPIELPVKEEPEKTVDVAAEKVKVITSEIPQTE 101
Db      225 EPKRTKRGVLGEPATPDKKENDAKSSDSVGEETLSPS--LKPEKVAEAEKKVEEAK 282
QY      102 RM---QKRAEFNVPVLSKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKERARQFG 158
Db      283 KKAEDQKEEDRRNYPTNTYKTLLEIAB--SDVEVK-----XAELELVKEEA---- 327
QY      159 LNVSSISRKSEDDKLRKRFIVTSSAGTGTEDTEAKRKRAE 205
Db      328 -----KEPRNEKVKQAKAE---VESQAEATRLNIKTDRKAE 364

RESULT 9
US-09-270-767-41091
; Sequence 41091, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41091
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-41091

Query Match      11.0%; Score 114.5; DB 4; Length 253;
Best Local Similarity 23.8%; Pred. No. 0.0058;
Matches 55; Conservative 35; Mismatches 80; Indels 61; Gaps 9;

QY      4 ETVELHKLKL---AELKQECLEAR-----GLETGK-----IKQDLIHLRLQAYLEEH 45
Db      55 EDIEVTKVAVLTDALFKACGGRTAHKGARHGLKLSGKIARLEQQEREMLEKLQRKLKTT 114
QY      46 ABEANEEDVLGDETEETEETKPIELPVKEEPEKTVDVAAEKVKVITSEIPQTERMQK 105
Db      115 PETEHRAVEVDSDSVEQCDDTK-----AKKKKKSIAEEFSEACEVSQ 155
QY      106 RAERFNPVLSKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKERARQFGLNYSIS 165
Db      156 LEE-----PIKSKKKKKDKAEGKESPAHQTPEDVPVKIKKKNKTRPGEDGVEV---- 207

RESULT 10
US-09-270-767-56307
; Sequence 56307, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56307
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-56307

Query Match      11.0%; Score 114.5; DB 4; Length 253;
Best Local Similarity 23.8%; Pred. No. 0.0058;
Matches 55; Conservative 35; Mismatches 80; Indels 61; Gaps 9;

QY      4 ETVELHKLKL---AELKQECLEAR-----GLETGK-----IKQDLIHLRLQAYLEEH 45
Db      55 EDIEVTKVAVLTDALFKACGGRTAHKGARHGLKLSGKIARLEQQEREMLEKLQRKLKTT 114
QY      46 ABEANEEDVLGDETEETEETKPIELPVKEEPEKTVDVAAEKVKVITSEIPQTERMQK 105
Db      115 PETEHRAVEVDSDSVEQCDDTK-----AKKKKKSIAEEFSEACEVSQ 155
QY      106 RAERFNPVLSKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKERARQFGLNYSIS 165
Db      156 LEE-----PIKSKKKKKDKAEGKESPAHQTPEDVPVKIKKKNKTRPGEDGVEV---- 207

RESULT 11
US-08-929-329-5
; Sequence 5, Application US/08929329
; Patent No. 6120770
; GENERAL INFORMATION:
; APPLICANT: Adams, John H
; APPLICANT: Dalton, John P
; APPLICANT: Kappe, Stefan
; TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing
; TITLE OF INVENTION: Vaccine Compositions
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 S Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/929,329
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 835910-28685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
```

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1507 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
;
US-08-929-329-5

Query Match 11.0%; Score 114.5; DB 3; Length 1507;
Best Local Similarity 24.9%; Pred. NO. 0.059;
Matches 52; Conservative 41; Mismatches 85; Indels 31; Gaps 7;

QY 2 ATETVELHLKLAEKQCLARGLETGKIQDLIHRLQAYLEEHAAEEANESDVLGDETE 61
DB 1114 AKKAEERKKAEAVKKAEEERKRRIEAEKKAEEERKRIEA--BKKAEE--RKRIEAEKKA 1169

QY 62 EEETKPIELPVKBEPEPEKTVDAAEKKVKVITTSIPQTERMQKRAERFNVPSLESKKA 121
DB 1170 EERKIIEAAKKAEEERKRRIEAEKKAEEERKKIEAAKKAEEERKKAEA--VKKAEEAKKK 1227

QY 122 ARAARFGISSVPTKGLSDNKPWNLDKLKRAQRFGLNVSSISRKSEDEK-----LKKR 177
DB 1228 AEAAK---KALKKXESRRGKKTIEAVKKAEEERKRI-----EAKKAEEERKRIEAVKKQ 1279

QY 178 KERFGIVTSSAGTGTTEDEAKKKRAER 206
DB 1280 KKK-----ENVEAEKKAEER 1295

RESULT 12
US-09-949-016-10326
; Sequence 10326, Application US/09949016
; Patent NO. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10326
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Human
;
US-09-949-016-10326

Query Match 10.9%; Score 113.5; DB 4; Length 1089;
Best Local Similarity 29.4%; Pred. NO. 0.048;
Matches 55; Conservative 32; Mismatches 77; Indels 23; Gaps 9;

QY 2 ATETVELHLKL--AELKQECIARGLETGKIQDLI--IHLQAYLEEHAAEE-----ANEE 53
DB 342 ASHTKPGKGLQVLKAKLQEAAMKLRFRERQKQKALFKLDNEDGFESEEESEEMTDESEE 401

QY 54 DVLGDETEEEETKPIELPVKBEPEPEKTVDAAEKKVKVITSEIPQTERMQKRAERFNV 113
DB 402 D--GEEKVKEKEKEEELSEEEKEEKEEENGGNQETAFLSSSEET-KDEKMDKENND 458

QY 114 VLSKKAARAARFGISSVPTKGLSSDNKPMVNDKLKERAQRFGLNVSSISRKSEDDSK 173

```


GenCore version 5.1.6
 Copyright (c) 1993 - 2005 Compugen Ltd.
 OW nucleic - nucleic search, using sw model
 Run on: October 4, 2005, 19:19:52 ; Search time 4044.85 Seconds
 (without alignments)
 10458.088 Million cell updates/sec

Title: US-09-788-476A-3
 Perfect score: 873
 Sequence: 1 tggagtggagggttaacaaga.....tgattatgttgcgttgtaaat 873

Scoring table: IDENTITY NUC
 Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
 2: gb_hcg.*
 3: gb_in.*
 4: gb_on.*
 5: gb_ov.*
 6: gb_pat.*
 7: gb_ph.*
 8: gb_pl.*
 9: gb_pr.*
 10: gb_ro.*
 11: gb_sts.*
 12: gb_sy.*
 13: gb_un.*
 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	873	100.0	1071	6	BD276528 FULL-LENG
2	873	100.0	1071	6	AX048092 Sequence
3	873	100.0	1520	6	AX405619 Sequence
4	873	100.0	1557	9	HSA409089
5	871.4	99.8	923	9	BC007099 Homo sapi
6	870.8	79.1	2553	6	CQ413602
7	647.4	74.2	84113	9	AC073063
8	633.6	72.6	189507	9	AF486281
9	633	72.5	633	9	AC145981
10	622.4	71.3	220280	2	AC145981
11	599.6	68.7	155666	10	AC134830
12	599.6	68.7	159681	10	AC132088
13	595.2	68.2	2553	6	CQ413602
14	586.2	67.1	908	10	BC027510
15	540.4	61.9	148152	2	AC022200 Homo sapi
16	540.4	61.9	149312	2	AC073328 Homo sapi
17	540.4	61.9	153936	9	AF001207 Homo sapi
18	540.4	61.9	166471	2	AC091052 Homo sapi
19	525	60.1	553	6	AR415990 Sequence

20	525	60.1	553	6	AX972824	Sequence
21	525	60.1	553	6	BD111543	BD111543 EST and e
22	469.2	53.7	539	11	BV179620	BV179620 sqnm10607
c 23	466.2	53.4	254288	2	AC126112	AC126112 Rattus no
24	465.6	53.3	471	6	AX884873	AX884873 Sequence
25	465.6	53.3	471	6	BD024483	BD024483 Sequence
26	458.4	52.5	488	6	CQ684732	CQ684732 Sequence
27	455	52.1	470	6	CQ683930	CQ683930 Sequence
28	431	49.4	470	6	CQ695331	CQ695331 Sequence
29	422	48.3	887	5	AJ719689	AJ719689 Gallus ga
30	415.8	47.6	643	5	BX932722	BX932722 Gallus ga
31	386	44.2	426	6	CQ695645	CQ695645 Sequence
32	379	43.4	379	6	CQ698322	CQ698322 Sequence
33	374.2	42.9	405	6	AR427142	AR427142 Sequence
34	374.2	42.9	405	6	AX987836	AX987836 Sequence
35	374.2	42.9	405	6	BD122695	BD122695 EST and e
36	334	38.3	408	6	AX210422	AX210422 Sequence
37	331.6	38.0	383	6	AR427141	AR427141 Sequence
38	331.6	38.0	383	6	AX987835	AX987835 Sequence
39	331.6	38.0	383	6	BD122694	BD122694 EST and e
40	307.2	35.2	797	5	CR760849	CR760849 Xenopus t
c 41	265.4	30.4	54666	9	AC073487	AC073487 Homo sapi
c 42	265.4	30.4	105736	2	AC090678	AC090678 Homo sapi
43	258.6	29.6	266	6	BD071528	BD071528 Secreted
44	255.8	29.3	204050	10	AL672298	AL672298 Mouse DNA
45	241	27.6	337	6	AR426739	AR426739 Sequence

ALIGNMENTS

RESULT 1
 BD276528 1071 bp DNA linear PAT 17-JUL-2003
 LOCUS FULL-LENGTH MOLECULES EXPRESSED IN HUMAN TISSUES.

BD276528

BD276528

BD276528.1 GI:33086296

JP 2002543839-A/31.

Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Shih,L.L., AuMyoung,J.L., Lu,D.A.M., Lal,P., Batra,S., Tang,T.Y.,

Yang,J., Azimzai,Y., Reddy,R., Henry, Yue and Baughn,M.R.

FULL-LENGTH MOLECULES EXPRESSED IN HUMAN TISSUES

Patent: JP 2002543839-A 31 24-DEC-2002;

INCYTE GENOMICS INC,Henry YUE,Tom Y TANG,Preeti IAL,Roopa REDDY,

Sajeew BAIRA,Mariah R BAUGHN,Jumming YANG,falda AZIMZAI, Dyung Aina

M LU, Janice AU-YOUNG,Leo L SHIH

OS Homo sapiens

PN JP 2002543839-A/31

PD 24-DEC-2002

PF 12-MAY-2000 JP 2000618453

PR 14-MAY-1999 US 09/311937,14-MAY-1999 US 08/311940, PR

PI 14-MAY-1999 US 09/311894

m lu,preeti lal,sajeew

PI batra,

PI tom y tang,jumming yang,yalda azimzai,roopa reddy,henry PI

Yue,mariah r baughn

CC This description about <220> can't be interpreted CC <220>

CC <221> misc feature

CC <223> Incyte ID No: 4093555CB1

Key Location/Qualifiers.

1. .1071

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

FEATURES

source

ORIGIN

Query Match 100.0%; Score 873; DB 6; Length 1071;

Best Local Similarity 100.0%; Pred. No. 8.2e-205; Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	TGAGTGGGGTAA	CAAGATGCGC	AGCGTGGAG	CTCCATA	AGCTAAAG	CTTGC	60	
Db	165	TGAGTGGGGTAA	CAAGATGCGC	AGCGTGGAG	CTCCATA	AGCTAAAG	CTTGC	224	
Qy	61	CGAAGTAAAG	CAAGATGCT	CTGCTGCT	TTGGAG	CAAGGG	ATTAAGCA	120	
Db	225	CGAAGTAAAG	CAAGATGCT	CTGCTGCT	TTGGAG	CAAGGG	ATTAAGCA	284	
Qy	121	TATCCACAG	ACTCCAGG	CAATCTT	GAAGAA	CACTGCT	GAAGAG	180	
Db	285	TATCCACAG	ACTCCAGG	CAATCTT	GAAGAA	CACTGCT	GAAGAG	344	
Qy	181	TGACTGGG	AGATGAAC	AGAGAA	GAAGAA	CAAA	CAAGTTC	240	
Db	345	TGACTGGG	AGATGAAC	AGAGAA	GAAGAA	CAAA	CAAGTTC	404	
Qy	241	GGAGAACCC	CTGAAA	AACTGTT	GGAG	CAAG	GGATTAAG	300	
Db	405	GGAGAACCC	CTGAAA	AACTGTT	GGAG	CAAG	GGATTAAG	464	
Qy	301	ATCTGAAT	ACCAAG	ACTGAG	AAATG	CGAG	GGCTGA	360	
Db	465	ATCTGAAT	ACCAAG	ACTGAG	AAATG	CGAG	GGCTGA	524	
Qy	361	GAGCTGG	AGATGA	AACTG	CTGAG	AACTG	CTGAG	420	
Db	525	GAGCTGG	AGATGA	AACTG	CTGAG	AACTG	CTGAG	584	
Qy	421	AAAAGTCT	GCATG	ATAC	CAAA	CACTAT	GGTTAA	480	
Db	585	AAAAGTCT	GCATG	ATAC	CAAA	CACTAT	GGTTAA	644	
Qy	481	TCAAAGAT	TTGGTT	GAATG	CTCTT	CAATCT	CCAGAA	540	
Db	645	TCAAAGAT	TTGGTT	GAATG	CTCTT	CAATCT	CCAGAA	704	
Qy	541	GAAGAG	AGAGAG	CGATG	TTGG	GAATG	TTGG	600	
Db	705	GAAGAG	AGAGAG	CGATG	TTGG	GAATG	TTGG	764	
Qy	601	GGATAC	AGAGCA	AGAGCA	AGAGCA	AGAGCA	AGAGCA	660	
Db	765	GGATAC	AGAGCA	AGAGCA	AGAGCA	AGAGCA	AGAGCA	824	
Qy	661	TCCTGAT	CTCTT	CTCTT	CTCTT	CTCTT	CTCTT	720	
Db	825	TCCTGAT	CTCTT	CTCTT	CTCTT	CTCTT	CTCTT	884	
Qy	721	TGCTAA	ATGCAC	AGTCA	TGCT	CGCA	ATGAGG	780	
Db	885	TGCTAA	ATGCAC	AGTCA	TGCT	CGCA	ATGAGG	944	
Qy	781	GGTACAT	CCATG	ATGTT	GAATG	TTGCT	TTGCT	840	
Db	945	GGTACAT	CCATG	ATGTT	GAATG	TTGCT	TTGCT	1004	
Qy	841	GTGTTT	TTGTTT	GAATG	TTGCT	TTGCT	TTGCT	873	
Db	1005	GTGTTT	TTGTTT	GAATG	TTGCT	TTGCT	TTGCT	1037	

RESULT 2
AX048092
LOCUS AX048092 1071 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 86 from Patent WO0070047.
ACCESSION AX048092
VERSION AX048092.1 GI:11876915
KEYWORDS Homo sapiens (human)
SOURCE ORGANISM Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 Yue,H., Tang,Y.T., Lal,P., Reddy,R., Batra,S., Baughn,M.R., Yang,J., Azimzai,Y., Lu,D.A., Au-Young,J. and Shih,L.L.
Full-length molecules expressed in human tissues
Patent: WO 0070047-A 86 23-NOV-2000;
Incyte Genomics, Inc. (US)
Location/Qualifiers
source 1..1071
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 4093555CB1"

ORIGIN

Query Match 100.0%; Score 873; DB 6; Length 1071; Best Local Similarity 100.0%; Pred. No. 8.2e-205; Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	TGAGTGGGGTAA	CAAGATGCGC	AGCGTGGAG	CTCCATA	AGCTAAAG	CTTGC	60	
Db	165	TGAGTGGGGTAA	CAAGATGCGC	AGCGTGGAG	CTCCATA	AGCTAAAG	CTTGC	224	
Qy	61	CGAAGTAAAG	CAAGATGCT	CTGCTGCT	TTGGAG	CAAGGG	ATTAAGCA	120	
Db	225	CGAAGTAAAG	CAAGATGCT	CTGCTGCT	TTGGAG	CAAGGG	ATTAAGCA	284	
Qy	121	TATCCACAG	ACTCCAGG	CAATCTT	GAAGAA	CACTGCT	GAAGAG	180	
Db	285	TATCCACAG	ACTCCAGG	CAATCTT	GAAGAA	CACTGCT	GAAGAG	344	
Qy	181	TGACTGGG	AGATGAAC	AGAGAA	GAAGAA	CAAA	CAAGTTC	240	
Db	345	TGACTGGG	AGATGAAC	AGAGAA	GAAGAA	CAAA	CAAGTTC	404	
Qy	241	GGAGAACCC	CTGAAA	AACTGTT	GGAG	CAAG	GGATTAAG	300	
Db	405	GGAGAACCC	CTGAAA	AACTGTT	GGAG	CAAG	GGATTAAG	464	
Qy	301	ATCTGAAT	ACCAAG	ACTGAG	AAATG	CGAG	GGCTGA	360	
Db	465	ATCTGAAT	ACCAAG	ACTGAG	AAATG	CGAG	GGCTGA	524	
Qy	361	GAGCTGG	AGATGA	AACTG	CTGAG	AACTG	CTGAG	420	
Db	525	GAGCTGG	AGATGA	AACTG	CTGAG	AACTG	CTGAG	584	
Qy	421	AAAAGTCT	GCATG	ATAC	CAAA	CACTAT	GGTTAA	480	
Db	585	AAAAGTCT	GCATG	ATAC	CAAA	CACTAT	GGTTAA	644	
Qy	481	TCAAAGAT	TTGGTT	GAATG	CTCTT	CAATCT	CCAGAA	540	
Db	645	TCAAAGAT	TTGGTT	GAATG	CTCTT	CAATCT	CCAGAA	704	
Qy	541	GAAGAG	AGAGAG	CGATG	TTGG	GAATG	TTGG	600	
Db	705	GAAGAG	AGAGAG	CGATG	TTGG	GAATG	TTGG	764	
Qy	601	GGATAC	AGAGCA	AGAGCA	AGAGCA	AGAGCA	AGAGCA	660	
Db	765	GGATAC	AGAGCA	AGAGCA	AGAGCA	AGAGCA	AGAGCA	824	
Qy	661	TCCTGAT	CTCTT	CTCTT	CTCTT	CTCTT	CTCTT	720	
Db	825	TCCTGAT	CTCTT	CTCTT	CTCTT	CTCTT	CTCTT	884	
Qy	721	TGCTAA	ATGCAC	AGTCA	TGCT	CGCA	ATGAGG	780	
Db	885	TGCTAA	ATGCAC	AGTCA	TGCT	CGCA	ATGAGG	944	
Qy	781	GGTACAT	CCATG	ATGTT	GAATG	TTGCT	TTGCT	840	
Db	945	GGTACAT	CCATG	ATGTT	GAATG	TTGCT	TTGCT	1004	
Qy	841	GTGTTT	TTGTTT	GAATG	TTGCT	TTGCT	TTGCT	873	
Db	1005	GTGTTT	TTGTTT	GAATG	TTGCT	TTGCT	TTGCT	1037	

```
|||||
945 GGATACATCCATGAACTCGGAGAGATTTGACATTAATGCTGTTTTCAGCTTTAAAGTTGTT 1004
Qy 841 GTGTTTTGTTTTGATTATGTTGCTTGTAAAT 873
Db 1005 GTGTTTTGTTTTGATTATGTTGCTTGTAAAT 1037

RESULT 3
AX405619
LOCUS AX405619 1520 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 34 from Patent WO0222660.
ACCESSION AX405619
VERSION AX405619.1 GI:21438677
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1
Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,
Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 022660-A 34 21-MAR-2002;
HYSEQ, INC. (US)
FEATURES
source Location/Qualifiers
1..1520
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
636..1268
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD34726.1"
/db_xref="GI:21438678"
/translation="MATETVELHKLKLAELKQELARGLTKGIKODLIHRLQALYLEE
HABEENEEDVLDGETTEETKEPIELPVKSEPEPEKTVDAAEKKVKITSEIPQTER
MQRARFNVFVLESKKAARAFGISVPTKGLSSDNKPMVNLDKKRAORFGLN
VSSISRKSEDEDEKKRERFVITSSAGTGTTEDEAKKRKAERFGIA"
ORIGIN
Query Match 100.0%; Score 873; DB 6; Length 1520;
Best Local Similarity 100.0%; Pred. NO. 8.3e-205;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TCGAGTGAAGGTTAAAGATGGCGACCGAGCGTGGAGCTCCATAGCTTAAAGCTTGC 60
Db 617 TGGAGTGAAGGTTAAAGATGGCGACCGAGCGTGGAGCTCCATAGCTTAAAGCTTGC 676
Qy 61 CGAAGTAAAGCAAGATGTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCT 120
Db 677 CGAAGTAAAGCAAGATGTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCT 736
Qy 121 TATCCACAGACTCAGGCATATCTTGAAGACATGCTGAAGAGGAGGCAATGAAGAAGA 180
Db 737 TATCCACAGACTCAGGCATATCTTGAAGACATGCTGAAGAGGAGGCAATGAAGAAGA 796
Qy 181 TGTAAGTGAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
Db 797 TGTAAGTGAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
Qy 241 GGAAGAACCCCTGAAAAAAGTGTGTATGTTGGCAGCAGAGAGGAGGAGGAGGAGGAGGAG 300
Db 857 GGAAGAACCCCTGAAAAAAGTGTGTATGTTGGCAGCAGAGAGGAGGAGGAGGAGGAGGAG 916
Qy 301 ATCTGAATACACAGACTGAGAGAAATGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
Db 917 ATCTGAATACACAGACTGAGAGAAATGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 976
Qy 361 GAGCTTGAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
Db 977 GAGCTTGAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1036
```

```
Qy 421 AAAAGGCTCTGATCTGATAAACAACCTATGTTTACTGTAAGCTGAAGGAAGAGC 480
Db 1037 AAAAGGCTCTGATCTGATAAACAACCTATGTTTAACTTGGATAAGCTGAAGGAAGAGC 1096
Qy 481 TCAAAGATTTGGTTTGAATGCTCTTCAATCTCAGAAAAGTCTGAAGATGATGAGAACT 540
Db 1097 TCAAAGATTTGGTTTGAATGCTCTTCAATCTCAGAAAAGTCTGAAGATGATGAGAACT 1156
Qy 541 GAAAAAGAGGAAGAGGAGGATTTGGGATTTGCACAAGTTTCAGCTGGAACTGGAACACAGA 600
Db 1157 GAAAAAGAGGAAGAGGAGGATTTGGGATTTGCACAAGTTTCAGCTGGAACTGGAACACAGA 1216
Qy 601 GGATACAGAGGCAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
Db 1217 GGATACAGAGGCAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1276
Qy 661 TCCTGATACATTTCTGTTTCTCCAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 1277 TCCTGATACATTTCTGTTTCTCCAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCT 1336
Qy 721 TGCCTAAATGCAACAGTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 1337 TGCCTAAATGCAACAGTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1396
Qy 781 GGTACATCCATGAACCTGCGGAGCAGCTTTCAGCTTATGCTGCTTTCAGCTTTAAGGTTGTT 840
Db 1397 GGTACATCCATGAACCTGCGGAGCAGCTTTCAGCTTATGCTGCTTTCAGCTTTAAGGTTGTT 1456
Qy 841 GTGTTTTGTTTTGATTATGTTGCTTGTAAAT 873
Db 1457 GTGTTTTGTTTTGATTATGTTGCTTGTAAAT 1489

RESULT 4
HSA409089
LOCUS HSA409089 1557 bp DNA linear PRI 11-MAY-2001
DEFINITION Homo sapiens HCC-1 gene.
ACCESSION AJ409089
VERSION AJ409089.1 GI:13940309
KEYWORDS HCC-1 gene; hepatocellular carcinoma; nuclear protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1
Choong, M.L., Tan, L.K., Lo, S.L., Ren, E.C., Ou, K., Ong, S.E.,
Liang, R.C., Sew, T.K. and Chung, M.C.
TITLE An integrated approach in the discovery and characterization of a
novel nuclear protein over-expressed in liver and pancreatic tumors
JOURNAL FEBS Lett. 496 (2-3), 109-116 (2001)
MEDLINE 21255661
PUBMED 11356193
REFERENCE 2 (bases 1 to 1557)
AUTHORS Choong, M.L.
DIRECT SUBMISSION
TITLE Submitted (21-FEB-2001) Choong M.L., National University of
Singapore, Bioprocessing Technology Centre, MD 11 Level 5, 10
Medical Drive, 117597, SINGAPORE
FEATURES
source Location/Qualifiers
1..1557
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q22"
/cell_line="HCC-M"
/cell_type="hepatocellular carcinoma"
/tissue_type="liver"
1..1557
/gene="HCC-1"
1..687
/gene="HCC-1"
/evidence="experimental"
promoter
```


Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IMAGE Plate: 21 Row: 9 Column: 18
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21361588.

FEATURES

Location/Qualifiers
1..923
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:14726 IMAGE:4273903"
/tissue_type="Prostate"
/clone_lib="NIH MGC_83"
/lab_host="DH10B"
/lab_host="pDNR-LIB"
/note="vector: pDNR-LIB"
1..923
/gene="CIP29"
/note="synonyms: MGC14726, HCC-1, HSPC316"
/db_xref="LocusID:84324"
22..654
/gene="CIP29"
/codon_start=1
/product="cytokine induced protein 29 kDa"
/protein_id="AAH07099.1"
/db_xref="GI:13937971"
/db_xref="LocusID:84324"
/translation="MATVELHLKUALKQECILARGLFTKGIKQDLIHLRLQYLBE
HAEEANEEDVLGDETEETKPIELPVKEEPEKTVDAAEKKVVKITSEIPQTER
MOKAERFNVPVLESKAAARAFGISVPTKGLSDNKPWNLDLKKRAORFGLN
VSSISRSKEDDEKLKKRERFGIVTSSAGTGTEDTEAKKKRAERFGIA"

gene

CDS

ORIGIN

Query Match 99.8%; Score 871.4; DB 9; Length 923;
Best Local Similarity 99.9%; Pred. NO. 2e-204;
Matches 872; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGAGTGGGGTAAACAGATGGGACCGAGACGGTGGAGCTCCATAGCTAAAGCTTGC 60
Db 3 TGGAGTGGGGTAAACAGATGGGACCGAGACGGTGGAGCTCCATAGCTAAAGCTTGC 62

Qy 61 CGAACTAAAGCAAGATCTCTTGTCTGCTGCTTGGAGACCAAGGAATAAAGCAAGATCT 120
Db 63 CGAACTAAAGCAAGATCTCTTGTCTGCTGCTTGGAGACCAAGGAATAAAGCAAGATCT 122

Qy 121 TATCCACAGACTCCAGGCATATCTTTGAAGAACATGCTGTAAGAGGAGGCAAAATGAAGA 180
Db 123 TATCCACAGACTCCAGGCATATCTTTGAAGAACATGCTGTAAGAGGAGGCAAAATGAAGA 182

Qy 181 TGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTAGCTCCCTGTCAGAGA 240
Db 183 TGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTAGCTCCCTGTCAGAGA 242

Qy 241 GGAAGACCCCTGAAAACATGTTGATGTGCGACGACAGAGGAAGGTGGAATTTAC 300
Db 243 GGAAGACCCCTGAAAACATGTTGATGTGCGACGACAGAGGAAGGTGGAATTTAC 302

Qy 301 ATCTGAATATCCACAGACTGAGAGAAATGCAGAAAGGCGCTGAACATTTCAATGTACCTGT 360
Db 303 ATCTGAATATCCACAGACTGAGAGAAATGCAGAAAGGCGCTGAACATTTCAATGTACCTGT 362

Qy 361 GAGCTTGAGAGTGAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAC 420
Db 363 GAGCTTGAGAGTGAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAC 422

Qy 421 AAAAGGTCTGTCTGATACAAACCTATGGTTAACTTGGATAGCTGAGAGGAAGGC 480
Db 423 AAAAGGTCTGTCTGATACAAACCTATGGTTAACTTGGATAGCTGAGAGGAAGGC 482

Qy 481 TCAGAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGCTCTGAAGATGATGAGAACT 540
Db 483 TCAGAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGCTCTGAAGATGATGAGAACT 542

Qy 541 GAAAAGAGGAAGAGGAGGATTTGGGATTTGTCAAAAGTTCAAGCTGGAAGTGAACACACAGA 600
Db 543 GAAAAGAGGAAGAGGAGGATTTGGGATTTGTCAAAAGTTCAAGCTGGAAGTGAACACACAGA 602

Qy 601 GGATACAGAGCAAAAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCGGATTCGCTGATGAAAGT 660
Db 603 GGATACAGAGCAAAAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCGGATTCGCTGATGAAAGT 662

Qy 661 TCCTGATACCTTCTGTTCTCCAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 663 TCCTGATACCTTCTGTTCTCCAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 722

Qy 721 TGCCTAAATGCACAGTCATGTCCTACGCTGCTCGCTCGCAATGAGGAGCATGTACCCCA 780
Db 723 TGCCTAAATGCACAGTCATGTCCTACGCTGCTCGCTCGCAATGAGGAGCATGTACCCCA 782

Qy 781 GGATACATCCATGAACCTGCGGACGAGTTTGACTTATTTGCTGTTTCAAGCTTTAAAGTGT 840
Db 783 GGATACATCCATGAACCTGCGGACGAGTTTGACTTATTTGCTGTTTCAAGCTTTAAAGTGT 842

Qy 841 GTGTTTTTGTGTTTGTGATTATGTTGCTTTGTTAAT 873
Db 843 GTGTTTTTGTGTTTGTGATTATGTTGCTTTGTTAAT 875

RESULT 6
CQ413602
LOCUS CQ413602 2553 bp DNA linear PAT 23-JAN-2004
DEFINITION Sequence 20673 from Patent WO0170979.
ACCESSION CQ413602
VERSION CQ413602.1 GI:41321383
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
AUTHORS Lee, J. and Lillie, J.
TITLE Genes, compositions, kits, and method for identification,
assessment, prevention, and therapy of ovarian cancer
JOURNAL Patent: WO 0170979-A 20673 27-SBP-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1..2553
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 79.1%; Score 690.8; DB 6; Length 2553;
Best Local Similarity 99.7%; Pred. No. 8.5e-160;
Matches 692; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 93 TGGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAC 152
Db 715 TGGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAC 774

Qy 153 ATGCTGAAGAGGAGGCGCAATGAAGAAGATGTACTTGGGAGATGAACAGAGGAAGAGAAA 212
Db 775 ATGCTGAAGAGGAGGCGCAATGAAGAAGATGTACTTGGGAGATGAACAGAGGAAGAGAAA 834

Qy 213 CAAAGCCCATTTAGCTCCCTGTCMAAGAGAGAACCCCTCGAAGAAAACCTGTTGATGTGG 272
Db 835 CAAAGCCCATTTAGCTCCCTGTCMAAGAGAGAACCCCTCGAAGAAAACCTGTTGATGTGG 894

Qy 273 CAGCAGAGAGAGAGTGGTGAATATTACATCTGAATATACACAGACTGAGAGAAATGCAGA 332
Db 895 CAGCAGAGAGAGAGTGGTGAATATTACATCTGAATATACACAGACTGAGAGAAATGCAGA 954


```
JOURNAL Biochem. Biophys. Res. Commun. 292 (3), 593-600 (2002)
MEDLINE 21920340
PUBMED 11922608
REFERENCE 2 (bases 1 to 633)
AUTHORS Fukuda,S., Wu,D.W., Stark,K. and Pelus,L.M.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Microbiology/Immunology, Indiana
University, School of Medicine, 1044 West Walnut Street,
Indianapolis, IN 46202-5121, USA
FEATURES
    source
        1..633
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /tissue_type="primary human breast cancer"
            1..633
                /note="nuclear protein with SAP motif"
                /codon_start=1
                /product="proliferation associated cytokine-inducible
                protein CIP29"
                /protein_id="AA09686.1"
                /db_xref="GI:20069116"
                /translation="MATETVELHLKLAEKQELARGLETGKIKDILHRLQYLVEE
                HAREANEEDVLGDTEETETKPIELPVKEEPEKTVDVAAEKVKVITSEIPOTER
                MQRAERNPVSVLESKARAARAFGISVPTKGLSDNKPWNLDLKLKERAQRFLGN
                VSSISRSKSEDDKLUKKRKRPGIVTSSAGTGTTEDTEAKKKRAERFGIA"
ORIGIN
Query Match 72.5%; Score 633; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.5e-145;
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 20 ATGCGCAGCGAGCGGTGGAGCTCATAAGCTAAAGCTTGCGAAGCTAAAGCAAGAATGT 79
Db 1 ATGCGCAGCGAGCGGTGGAGCTCATAAGCTAAAGCTTGCGAAGCTAAAGCAAGAATGT 60
Qy 80 CTTGCTGTGTTGGAGACCAAGGATTAAGCAAGATCTTATCCAGACTCCAGGCA 139
Db 61 CTTGCTGTGTTGGAGACCAAGGATTAAGCAAGATCTTATCCAGACTCCAGGCA 120
Qy 140 TATCTTGAAGAACATGCTGAAGAGGAGCAATGAAGAGATGTACTGGGAGATGAACA 199
Db 121 TATCTTGAAGAACATGCTGAAGAGGAGCAATGAAGAGATGTACTGGGAGATGAACA 180
Qy 200 GAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTGAAAAA 259
Db 181 GAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTGAAAAA 240
Qy 260 ACTGTTGATGCGCAGCAGAGAGAAAGTGGTGAATAATTAATCTGAAATACCAAGACT 319
Db 241 ACTGTTGATGCGCAGCAGAGAGAAAGTGGTGAATAATTAATCTGAAATACCAAGACT 300
Qy 320 GAGGAATGCAGAGAGGCGTGAACGATTCATCTGAGTTCAGCTTGAGAGTAAGAAA 379
Db 301 GAGGAATGCAGAGAGGCGTGAACGATTCATCTGAGTTCAGCTTGAGAGTAAGAAA 360
Qy 380 GCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCACAAAGAGTCTCTCATCTGAT 439
Db 361 GCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCACAAAGAGTCTCTCATCTGAT 420
Qy 440 AACAAACCTATGTTAACTTCGATAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAAT 499
Db 421 AACAAACCTATGTTAACTTCGATAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAAT 480
Qy 500 GTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAGAGAGAGGAGCGA 559
Db 481 GTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAGAGAGAGGAGCGA 540
Qy 560 TTTGGGATTGTACAAAGTTACAGTTGGAACCTGGAACCAAGAGGATACAGAGGCAAGAG 619
Db 541 TTTGGGATTGTACAAAGTTACAGTTGGAACCTGGAACCAAGAGGATACAGAGGCAAGAG 600
Qy 620 AGGAAAAGAGCAGAGCGCTTTGGGATTGCTGTA 652
```

```
Db 601 AGGAAAAGAGCAGAGCGCTTTGGGATTGCTGTA 633
|||||
AC145981 220280 bp DNA linear HTG 01-AUG-2003
Pan troglodytes chromosome UNK clone RP43-13G23, *** SEQUENCING IN
PROGRESS ***, 46 unordered pieces.
AC145981
AC145981.1 GI:33386932
HTG; HTGS PHASE1.
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Wilson,R.K.
REFERENCE
1 (bases 1 to 220280)
The sequence of Pan troglodytes clone
Unpublished
2 (bases 1 to 220280)
Wilson,R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: C.PT013G23
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1338: contig of 1338 bp in length
* 1339 1438: gap of unknown length
* 1439 2637: contig of 1199 bp in length
* 2638 2737: gap of unknown length
* 2738 4065: contig of 1328 bp in length
* 4066 4165: gap of unknown length
* 4166 5254: contig of 1089 bp in length
* 5255 5355: gap of unknown length
* 5355 7079: contig of 1724 bp in length
* 7079 8271: gap of unknown length
* 8271 8371: contig of 1093 bp in length
* 8371 9516: contig of 1145 bp in length
* 9516 9617: gap of unknown length
* 9617 10946: contig of 1329 bp in length
* 10946 12864: contig of 1819 bp in length
* 12864 12964: gap of unknown length
* 12965 15017: contig of 2053 bp in length
* 15018 15117: gap of unknown length
* 15118 16693: contig of 1576 bp in length
* 16694 16794: gap of unknown length
* 16794 18814: contig of 1921 bp in length
* 18815 21332: contig of 2518 bp in length
* 21333 21432: gap of unknown length
* 21433 23240: contig of 1808 bp in length
* 23241 23341: gap of unknown length
* 23341 24932: contig of 1592 bp in length
* 24933 25032: gap of unknown length
* 25033 27092: contig of 2060 bp in length
* 27092 27092: contig of 2060 bp in length
```

```
* 27093 27192: gap of unknown length
* 28193 28362: Contig of 1170 bp in length
* 28363 28462: gap of unknown length
* 28463 29890: Contig of 1428 bp in length
* 29891 29990: gap of unknown length
* 29991 31733: Contig of 1743 bp in length
* 31734 31833: gap of unknown length
* 31834 34959: Contig of 3126 bp in length
* 34960 35059: gap of unknown length
* 35060 37949: Contig of 2890 bp in length
* 37950 38049: gap of unknown length
* 38050 41435: Contig of 3386 bp in length
* 41436 41535: gap of unknown length
* 41536 46033: Contig of 4498 bp in length
* 46034 46133: gap of unknown length
* 46134 49523: Contig of 3390 bp in length
* 49524 49623: gap of unknown length
* 49624 54040: Contig of 4417 bp in length
* 54041 54140: gap of unknown length
* 54141 58313: Contig of 4173 bp in length
* 58314 58413: gap of unknown length
* 58414 62876: Contig of 4463 bp in length
* 62877 62976: gap of unknown length
* 62977 69574: Contig of 6598 bp in length
* 69575 69674: gap of unknown length
* 69675 75666: Contig of 5992 bp in length
* 75667 75766: gap of unknown length
* 75767 81362: Contig of 5596 bp in length
* 81363 81462: gap of unknown length
* 81463 85272: Contig of 3810 bp in length
* 85273 85372: gap of unknown length
* 85373 91974: Contig of 6602 bp in length
* 91975 92074: gap of unknown length
* 92075 96689: Contig of 4615 bp in length
* 96690 96790 103977: Contig of 7188 bp in length
* 103978 104077: gap of unknown length
* 104078 109657: Contig of 5580 bp in length
* 109658 109757: gap of unknown length
* 109758 116696: Contig of 6939 bp in length
* 116697 116796: gap of unknown length
* 116797 124010: Contig of 7214 bp in length
* 124011 124110: gap of unknown length
* 124111 132127: Contig of 8017 bp in length
* 132128 132227: gap of unknown length
* 132228 139538: Contig of 7311 bp in length
* 139539 139638: gap of unknown length
* 139639 148819: Contig of 9181 bp in length
* 148820 148919: gap of unknown length
* 148920 157172: Contig of 8253 bp in length
* 157173 157272: gap of unknown length
* 157273 168568: Contig of 11296 bp in length
* 168569 168668: gap of unknown length
* 168669 179527: Contig of 10859 bp in length
* 179528 179627: gap of unknown length
* 179628 188941: Contig of 9314 bp in length
* 188942 189041: gap of unknown length
* 189042 201122: Contig of 12081 bp in length
* 201123 201222: gap of unknown length
* 201223 220280: Contig of 19058 bp in length.
```

FEATURES

```
source
1. .220280
  /organism="Pan troglodytes"
  /mol_type="genomic DNA"
  /db_xref="taxon:9598"
  /chromosome="UNK"
  /clone="RP43-13G23"
  /note="assembly_name:Contig25"
misc_feature
1439. .2637
  /note="assembly_name:Contig33"
misc_feature
2738. .4065
  /note="assembly_name:Contig39"

Query Match 71.3%; Score 622.4; DB 2; Length 220280;
Best Local Similarity 88.5%; Pred.No. 7.6e-143;
Matches 768; Conservative 0; Mismatches 61; Indels 39; Gaps 7;

QY 7 GAGGGGTAACAAGATCGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTCCGCAACT 66
  |||||
Db 147627 GAGGGGTAACAAGATCGTAACTGAGATGGTGGAGCTCCGTAAGCTTGAAGCTTCTAACT 147686
  |||||

QY 67 AAAGCAAGAATGCTTCTCGTGGTGGAGACCAAGGGGAATAAGCAAGATCTTATCCA 126
  |||||
Db 147687 AAAGCAAGAATGCTTCTGTGTTGGAGACCAAGGGGAATAAGCAAGATCTTATCCA 147746
  |||||
```

```
Qy 127 CAGACTCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAGATGTACT 186
Db 147747 CAGACTCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAGATGTACT 147803
Qy 187 GGGAGATGAACACAGAGAAAGAAACAAAGCCCAATTGAGCTCCCTGCTCAAGAGGAGAGA 246
Db 147804 AGGAGATGAACACAGAGAAAGAAACAAAGCCCAATTGAGCTCCCTGCTCAAGAGGAGAGA 147850
Qy 247 ACCCCCTG-AAAACTGTTGATGTGGCAGCAGAGAGAAAGAGTGGTGAATTCATCTGAGCT 305
Db 147851 ACCCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGAGTGGTGAATTCATCTGAGCT 147910
Qy 306 AATATACACACATGAGAGATGAGAGAGGCTGACAGTCAATGCTACCTGAGCT 365
Db 147911 AATATACACACATGAGAGATGAGAGAGGCTGACAGTCAATGCTACCTGAGCT 147970
Qy 366 TGGAGAGTAAG- --AAAGCTCTCGGGCAGCTAGGTTGGGATTTCTTCAGTTCACAA 422
Db 147971 TGGAGAGTAAGTAAGTAATCAGGAGCTAGGTTGGGATTTCTTCAGTTCACAA 148030
Qy 423 AAGTCTGTCTATGATAACAACTATGGTTAACTTGGATAAGCTGAAGGAAAGAGCTC 482
Db 148031 AA-GTCTGTCTATG- --AACACACTATGGTTAACTTGGATAAGCGAAGGAAAGAGCTC 148087
Qy 483 AAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGCTGAGAGATGATGAGAACTGA 542
Db 148088 AAGATCTGGTTTGAATGCTCTTCAATCTCCAGAAAGCTGAGAGATGATAAGAACTGA 148147
Qy 543 AAAAGAGGAAGGAGCGATTTGGGATTTGCAAGATTCAGCTGGAAGTGGAAACACAGAG 602
Db 148148 AAAAGAGGAAGGAGCGATTTGGGATTTGCAAGATTCAGCTGGAAGTGGAAACACAGAG 148207
Qy 603 ATACAGAGGCAAGAGAAAGAGGAAAGAGCAGAGCGCTTTGGGATTCCTGATGAAAAGTTC 662
Db 148208 ATACAGAGGCAAGAGAAAGAGGAAAGAGCAGAGCGCTTTGGGATTCCTGATGAAAAGTTC 148267
Qy 663 CTGATATCTTCTGTTCCAGTGTTCCTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 722
Db 148268 TTGATATCTTCTGTTCTCCAGTGTTCCTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 148327
Qy 723 CCTAAATGCACAGTCATGTCGCTACGTCCTCGCTCGCAATGAGGAGGAGCATGCCCCAGG 782
Db 148328 CTAAA- -----TGAGCTCTTCCCTTGCATGAGGAGGAGCATGCCCCAGG 148371
Qy 783 TACATCCATGAACGCGCAGCAGTTTGACTTTATGCTGTTTTCAGCTTTAAGGTTGTGT 842
Db 148372 TAAACTGTGAACGCGCGCAGCTTTGACTTTATGCTGTTTTCAGCTTTAAGGTTGTGT 148431
Qy 843 GTTTTGTGTTTGAATGATGTTGCTGTT 870
Db 148432 GTTTTGTGTTTGAATGATGTTGCTGTT 148459

RESULT 11
AC134830 155666 bp DNA linear ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-498G16 from chromosome 19, complete
sequence.
ACCESSION AC134830
VERSION AC134830.2 GI:34495140
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
VanBrunt,A.
REFERENCE 1 (bases 1 to 155666)
TITLE The sequence of Mus musculus BAC clone RP24-498G16
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 155666)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
```

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

3 (bases 1 to 155666)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (01-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 155666)
Wilson,R.K.
Direct Submission
Submitted (07-SEP-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

5 (bases 1 to 155666)
Wilson,R.
Direct Submission
Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 7, 2003 this sequence version replaced gi:23396352.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics

Center project name: M_BB0498G16

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

FEATURES

source
Location/Qualifiers
1..155666
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="19"
/map="19"
/clone="RP24-498G16"
/clone_lib="RPCI-24"
1..118
/rpt_family="MaLR"
136..410
/rpt_family="MaLR"
472..670
/rpt_family="ERV1"
1983..2092
/rpt_family="MER2_type"
2103..2222
/rpt_family="BC1_MM"
2242..2546
/rpt_family="MER2_type"

repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

repeat_region	4917.4984	/rpt_family="L1"
repeat_region	5489.5558	/rpt_family="MaLR"
repeat_region	5559.5886	/rpt_family="MaLR"
repeat_region	5887.6640	/rpt_family="MaLR"
repeat_region	6679.6727	/rpt_family="MaLR"
repeat_region	6985.7169	/rpt_family="B2"
repeat_region	9274.9616	/rpt_family="MaLR"
repeat_region	9949.10089	/rpt_family="B4"
repeat_region	11280.11443	/rpt_family="ERV1"
repeat_region	11691.12093	/rpt_family="MaLR"
repeat_region	12146.12311	/rpt_family="L1"
repeat_region	12312.12494	/rpt_family="ERVK"
repeat_region	13916.14409	/rpt_family="ERVJ"
repeat_region	14410.14743	/rpt_family="MaLR"
repeat_region	14766.14939	/rpt_family="B2"
repeat_region	15237.15387	/rpt_family="MTR"
repeat_region	16072.16182	/rpt_family="MaLR"
repeat_region	16183.16651	/rpt_family="RMER6B"
repeat_region	16652.16842	/rpt_family="MaLR"
repeat_region	17041.17109	/rpt_family="ERV1"
repeat_region	17202.17280	/rpt_family="B2"
repeat_region	17439.17674	/rpt_family="L1"
repeat_region	19312.19449	/rpt_family="Alu"
repeat_region	20160.20251	/rpt_family="ID"
repeat_region	20693.20835	/rpt_family="B4"
repeat_region	20849.21036	/rpt_family="B2"
repeat_region	21284.21740	/rpt_family="L1"
repeat_region	21741.22138	/rpt_family="MaLR"
repeat_region	22139.22242	/rpt_family="L1"
repeat_region	22243.22241	/rpt_family="B2"
repeat_region	22422.24053	/rpt_family="L1"
repeat_region	24110.24259	/rpt_family="L1"
repeat_region	24546.24762	/rpt_family="L1"
repeat_region	25008.25215	/rpt_family="MaLR"
repeat_region	26951.27056	/rpt_family="ERVJ"
repeat_region	28629.29006	/rpt_family="MaLR"
repeat_region	29259.29379	

```

Qy 481 TCAAGATTGGTTGAATGCTCTTCAATCTCAGAAAGTCTGAAGATGATGAGAACT 540
Db 27556 ACAGAGATTGGTTGAATGCTCTTCCATCTCTAGAAAGTCTGAGGATGATGAGAACT 27615
Qy 541 GAAAGAGAGGAGGAGCGATTGGGATTGTGCACAAAGTTCAGCTGGAACCTGGAACACAGAG 600
Db 27616 GAAGAACGAGAGAGATTGGGATTGTGCACAAAGTTCAGCTGGAACCTGGAACACAGAG 27675
Qy 601 GGATACAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 27676 GGATACAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 27735
Qy 661 TCCTGATACATCTCTCTCCAGTGTCTTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 27736 TCGTG---CTTCTGACCATAGTGTCTTCCATCTCTCTCTCTCTCTCTCTCTCTCTCT 27792
Qy 721 TGCCTAAATGACAGTATGCTGCTACGTC-----CTGCCCTCGCAATGAG 765
Db 27793 TACATACACACCCCTACCTAGTATCTATCTACATACAGCTCATGTGCTCAGTGAG 27852
Qy 766 GGAGCATGTACCCCGAGGTACATCCATGAACCTGCGGACAGCTTTGACTTATTTGCTGTTTC 825
Db 27853 GAAGCATGTACGTTGGTGATCATCTTGAATTCAGTAGCAGTTTATTTAACTGCTCTTCT 27912
Qy 826 AGCTTTAAGTGTGCTGCTTTGTTTGTGTTTGTATGTT 863
Db 27913 GCGTTTAAGATTGTTGCTGTTTGTGTTTGTGTTTAAATTT 27950

```

RESULT 12

```

AC132088/c
LOCUS AC132088 159681 bp DNA linear ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-410M8 from chromosome 19, complete
sequence.
ACCESSION AC132088
VERSION AC132088.4 GI:34740419
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 159681)
AUTHORS Harkins, R. and Bielicki, L.
TITLE The sequence of Mus musculus BAC clone RP24-410M8
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 159681)
AUTHORS Wilson, R.K.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 159681)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 159681)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 159681)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 159681)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Sep 16, 2003 this sequence version replaced gi:33187079.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC

```

Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.edu
 ----- Summary Statistics
 Center project name: M_BB0410M08

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC124557 and AC134830.

FEATURES	Location/Qualifiers
source	1..159681 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /chromosome="19" /map="19" /clone="RP24-410M8" /clone_lib="RPCI-24"
repeat_region	1..405 /rpt_family="ERVr"
repeat_region	504..738 /rpt_family="MIR"
repeat_region	1016..1148 /rpt_family="B4"
repeat_region	1264..1691 /rpt_family="MaLR"
repeat_region	1696..1839 /rpt_family="MER2_type"
repeat_region	1912..2118 /rpt_family="L1"
repeat_region	2160..2247 /rpt_family="MIR"
repeat_region	2643..2829 /rpt_family="B2"
repeat_region	2831..3011 /rpt_family="B2"
trna	complement(2934..3006) /note="Likely pseudogene (HMM SC=29.93 / Sec struct SC=-6.06)"
repeat_region	4669..4878 /rpt_family="B2"
repeat_region	5508..5655 /rpt_family="RMR1B"
repeat_region	5743..5839 /rpt_family="L1"
repeat_region	5924..6010 /rpt_family="RMR1B"

```
repeat_region 6496..6633
/rpt_family="RMER1B"
6779..6884
/rpt_family="MIR"
repeat_region 6885..7304
/rpt_family="RMER17C"
repeat_region 7305..7333
/rpt_family="MIR"
repeat_region 9080..9263
/rpt_family="L1"
repeat_region 9503..9905
/rpt_family="MaLR"
repeat_region 12380..12477
/rpt_family="Alu"
repeat_region 12721..13257
/rpt_family="MaLR"
repeat_region 14045..14634
/rpt_family="L1"
repeat_region 16118..16187
/rpt_family="MER1_type"
repeat_region 16297..16617
/rpt_family="MaLR"
misc_feature 18720..19476
/note="CpG island (GC=72.4, o/e=0.96, #CpGs=99)"
repeat_region 2008..20080
/rpt_family="B4"
repeat_region 2081..20114
/rpt_family="Alu"
repeat_region 20300..20423
/rpt_family="B2"
repeat_region 20495..21885
/rpt_family="L1"
repeat_region 22810..22761
/rpt_family="B4"
repeat_region 23159..23333
/rpt_family="B4"
repeat_region 23761..23895
/rpt_family="Alu"
repeat_region 25164..25479
/rpt_family="MaLR"
repeat_region 25575..25762
/rpt_family="L1"
repeat_region 25806..26115
/rpt_family="L1"
repeat_region 27147..27426
/rpt_family="B4"
repeat_region 28480..28600
/rpt_family="MaLR"
repeat_region 28853..29230
/rpt_family="MaLR"
repeat_region 30803..30905
/rpt_family="ERV1"
repeat_region 32644..32851
/rpt_family="MaLR"
repeat_region 33097..33313
/rpt_family="L1"
repeat_region 33590..33749
/rpt_family="L1"
repeat_region 33806..35437
/rpt_family="L1"
repeat_region 35438..35616
/rpt_family="B2"
repeat_region 35617..35720
/rpt_family="L1"
repeat_region 35721..36118
/rpt_family="MaLR"
repeat_region 36119..36575
/rpt_family="L1"
repeat_region 36823..37010
/rpt_family="B2"
repeat_region 37024..37166
/rpt_family="B4"
repeat_region 37608..37699
```

```
repeat_region /rpt_family="ID"
38410..38547
/rpt_family="Alu"
repeat_region 40185..40420
/rpt_family="L1"
repeat_region 40579..40657
/rpt_family="B2"
repeat_region 40693..40766
/rpt_family="ERV1"
repeat_region 40965..41155
/rpt_family="MaLR"
repeat_region 41156..41624

Query Match 68.7%; Score 599.6; DB 10; Length 159681;
Best Local Similarity 82.1%; Pred.No.3.2e-137; Indels 18; Gaps 2;
Matches 721; Conservative 0; Mismatches 139;

Qy 1 TGGAGTGAAGGGGTAAACAAGATGCCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30783 TGGACTGAGGCGCATCAAGATGCGCGCGGAGACGGTGGAGCTCCACAAGCTGAAGCTTGC 30724
Qy 61 CGAAGCTAAAGCAAGAAATGCTTCTGCTCGTGGTGGAGACCAAGGGAAATAAAGCAAGATCT 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30723 TGAAGCTAAAGCAAGAAATGCTTCTGCTCGTGGTGGAGACCAAGGGAAATAAAGCAAGATCT 30664
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGAGGCGCAATGAAGAAGA 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30663 TATAAATAGGCTGCGAGCCCTATCTTGAAGACCATGAGAGAGGAGGAGCAATGAAGAAGA 30604
Qy 181 TGTACTGGGAGATGAACACAGAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30603 TGTACTGGGAGATGAACCTGAGAGAGAGAACCAAGAGCCATAGACTGCTGTAAAGA 30544
Qy 241 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGAAAGTGGTGAATAATTAC 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30543 GGAAGAACCTCTCTGAAAAAGGCTGTTGATATGGCATCGGAAAGAGGTTGTTAAAGATTAC 30484
Qy 301 ATCTGAAATACACACAGACTGAGAGAGATGAGAGAGGCGTGAACGATTCAATCTACTCTGT 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30483 ATCTGGAATACCTCAGACTGAAGAAATGAGAGAGGCGTGAACGATTCAATCTACTCTGT 30424
Qy 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTTCAAC 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30423 AAGCTTGGAGAGTAAGAGGCTGCCCGGGCGGCTAGGTTTGGAAATTTCTTCAGTTTCAAC 30364
Qy 421 AAAAGTCTGTCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGAAAGAGC 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30363 AAAAGGTTTATCATCTGACACCAAGGCTATGGTTAACTTGGATAAGCTGAAGAAAGAGC 30304
Qy 481 TCAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACT 540
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30303 ACAGAGATTTGGTTTGAATGCTCTTCCATCTCTAGAAAGTCTGAGAGATGATGAGAGCT 30244
Qy 541 GAAAAAGAGAGAGGAGCGGATTTGGGATTTGTCAAGTTTCAAGTTTCAAGTTTGGAACTGGAACCCACAGA 600
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30243 GAAGAACCGAGAGAGAGATTTGGGATTTGACAGTTTGGAGTTGAGCTGGAACCTGGAACCCAGA 30184
Qy 601 GGATACAGAGGCAAGAAAGAGAGGAAAAAGAGCAGAGCGCTTTGGGATTTGCGTGTGAAAGT 660
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30183 GGATACAGAGGCAAGAAAGAGAGGAAAAAGAGCAGAGCGCTTTTGGAAATTTGCATAATGAAAGT 30124
Qy 661 TCCTGATACCTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30123 TCGTG---CTTTCTGCACCATAGTGTGTTTCCATTTCTTCAAAATTTTCTTGGTTTATATA 30067
Qy 721 TGCCTAAATGCACAGTCAATGTGCTACGTC-----CTGCCCTCGCAATGAG 765
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30066 TACATACACACCCCTACCTACGTATCTATCTACATACACAGTCAATGTGCTCTACAGTGAG 30007
Qy 766 GGAGCATGTACCCCGAGTACATCCATGAACCTGCGGAGCAGGTTTGAGCTTTATTTGCTGTTTC 825
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30006 GAAGCATGTACGTTTGGTATCATCTTCTGAAATTCAGTAGCAGTGTATTTTAACTGCTCTTCT 29947
```

FEATURES	SOURCE
1. The first two columns are labeled "FEATURES" and "SOURCE".	
2. The third column is labeled "FEATURES".	
3. The fourth column is labeled "SOURCE".	
4. The fifth column is labeled "FEATURES".	
5. The sixth column is labeled "SOURCE".	
6. The seventh column is labeled "FEATURES".	
7. The eighth column is labeled "SOURCE".	
8. The ninth column is labeled "FEATURES".	
9. The tenth column is labeled "SOURCE".	
10. The eleventh column is labeled "FEATURES".	
11. The twelfth column is labeled "SOURCE".	
12. The thirteenth column is labeled "FEATURES".	
13. The fourteenth column is labeled "SOURCE".	
14. The fifteenth column is labeled "FEATURES".	
15. The sixteenth column is labeled "SOURCE".	
16. The seventeenth column is labeled "FEATURES".	
17. The eighteenth column is labeled "SOURCE".	
18. The nineteenth column is labeled "FEATURES".	
19. The twentieth column is labeled "SOURCE".	
20. The twenty-first column is labeled "FEATURES".	
21. The twenty-second column is labeled "SOURCE".	
22. The twenty-third column is labeled "FEATURES".	
23. The twenty-fourth column is labeled "SOURCE".	
24. The twenty-fifth column is labeled "FEATURES".	
25. The twenty-sixth column is labeled "SOURCE".	
26. The twenty-seventh column is labeled "FEATURES".	
27. The twenty-eighth column is labeled "SOURCE".	
28. The twenty-ninth column is labeled "FEATURES".	
29. The thirtieth column is labeled "SOURCE".	
30. The thirty-first column is labeled "FEATURES".	
31. The thirty-second column is labeled "SOURCE".	
32. The thirty-third column is labeled "FEATURES".	
33. The thirty-fourth column is labeled "SOURCE".	
34. The thirty-fifth column is labeled "FEATURES".	
35. The thirty-sixth column is labeled "SOURCE".	
36. The thirty-seventh column is labeled "FEATURES".	
37. The thirty-eighth column is labeled "SOURCE".	
38. The thirty-ninth column is labeled "FEATURES".	
39. The fortieth column is labeled "SOURCE".	
40. The forty-first column is labeled "FEATURES".	
41. The forty-second column is labeled "SOURCE".	
42. The forty-third column is labeled "FEATURES".	
43. The forty-fourth column is labeled "SOURCE".	
44. The forty-fifth column is labeled "FEATURES".	
45. The forty-sixth column is labeled "SOURCE".	
46. The forty-seventh column is labeled "FEATURES".	
47. The forty-eighth column is labeled "SOURCE".	
48. The forty-ninth column is labeled "FEATURES".	
49. The fiftieth column is labeled "SOURCE".	
50. The fifty-first column is labeled "FEATURES".	
51. The fifty-second column is labeled "SOURCE".	
52. The fifty-third column is labeled "FEATURES".	
53. The fifty-fourth column is labeled "SOURCE".	
54. The fifty-fifth column is labeled "FEATURES".	
55. The fifty-sixth column is labeled "SOURCE".	
56. The fifty-seventh column is labeled "FEATURES".	
57. The fifty-eighth column is labeled "SOURCE".	
58. The fifty-ninth column is labeled "FEATURES".	
59. The sixtieth column is labeled "SOURCE".	
60. The sixty-first column is labeled "FEATURES".	
61. The sixty-second column is labeled "SOURCE".	
62. The sixty-third column is labeled "FEATURES".	
63. The sixty-fourth column is labeled "SOURCE".	
64. The sixty-fifth column is labeled "FEATURES".	
65. The sixty-sixth column is labeled "SOURCE".	
66. The sixty-seventh column is labeled "FEATURES".	
67. The sixty-eighth column is labeled "SOURCE".	
68. The sixty-ninth column is labeled "FEATURES".	
69. The seventieth column is labeled "SOURCE".	
70. The seventy-first column is labeled "FEATURES".	
71. The seventy-second column is labeled "SOURCE".	
72. The seventy-third column is labeled "FEATURES".	
73. The seventy-fourth column is labeled "SOURCE".	
74. The seventy-fifth column is labeled "FEATURES".	
75. The seventy-sixth column is labeled "SOURCE".	
76. The seventy-seventh column is labeled "FEATURES".	
77. The seventy-eighth column is labeled "SOURCE".	
78. The seventy-ninth column is labeled "FEATURES".	
79. The eightieth column is labeled "SOURCE".	
80. The eighty-first column is labeled "FEATURES".	
81. The eighty-second column is labeled "SOURCE".	
82. The eighty-third column is labeled "FEATURES".	
83. The eighty-fourth column is labeled "SOURCE".	
84. The eighty-fifth column is labeled "FEATURES".	
85. The eighty-sixth column is labeled "SOURCE".	
86. The eighty-seventh column is labeled "FEATURES".	
87. The eighty-eighth column is labeled "SOURCE".	
88. The eighty-ninth column is labeled "FEATURES".	
89. The ninetieth column is labeled "SOURCE".	
90. The ninety-first column is labeled "FEATURES".	
91. The ninety-second column is labeled "SOURCE".	
92. The ninety-third column is labeled "FEATURES".	
93. The ninety-fourth column is labeled "SOURCE".	
94. The ninety-fifth column is labeled "FEATURES".	
95. The ninety-sixth column is labeled "SOURCE".	
96. The ninety-seventh column is labeled "FEATURES".	
97. The ninety-eighth column is labeled "SOURCE".	
98. The ninety-ninth column is labeled "FEATURES".	
99. The hundredth column is labeled "SOURCE".	

		/mol_type="mRNA"	
		/db_xref="taxon:10090"	
		/clone="MGC:41057 IMAGE:1382354"	
		/tissue_type="Mammary gland, lactating mouse"	
		/clone_lib="Soares_mammary_gland_NMLMG"	
		/lab_host="DH10B"	
		/notes=Vector: pI7T3-Pac"	
gene		1..908	
		/gene="1110005A23Rik"	
		/product="RIKEN CDNA 1110005A23"	
		/protein_id="AAH27510.1"	
		/db_xref="GI:20380578"	
		/db_xref="LocustID:66118"	
		/db_xref="MGI:1913368"	
CDS		5..637	
		/gene="1110005A23Rik"	
		/codon_start=1	
		/product="RIKEN CDNA 1110005A23"	
		/protein_id="AAH27510.1"	
		/db_xref="GI:20380578"	
		/db_xref="LocustID:66118"	
		/db_xref="MGI:1913368"	
		/translation="MAEATVELHLKLAELKQBCLAGLGTETGKI KODLINRLQAYLEED HAEEARENEEDVLDETEEBEPKEI ELPFKEEPPPEKAVDMSAEKKVKVITSGIPQTER MOKRAEFNPVPVLESKKAARAARFGISVPTPKGLSDTKPMWNLDKLKERAQRFGLN VSSI SRKSSEDEXLKXKRERFGIVTS SAGTGTTEDTEAKKRKRAERFGIA"	
ORIGIN			
	Query Match	67.1%; Score 586.2; DB 10; Length 908;	
	Best Local Similarity	81.9%; Pred. No. 5.7e-134;	
	Matches	707; Conservative 0; Mismatches 138; Indels 18; Gaps 2;	
QY	16	CAAGATGGCCACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGCCGAACCTAAAGCAAGA	75
Db	1	CAAGATGGCCGCCGAGACGGTGGAGCTCCACAAGCTGAAGCTTGCTGAACCTAAAGCAAGA	60
QY	76	ATGCTTTGCTCGTGGTTTTGGAGACCAAGGGAATAAAGCAAGATCTTTATCACACACTCCA	135
Db	61	ATGCTTTGCTCGTGGTTTTAGAGACCAGGGAATAAAAACAGATCTTTATAATAGGCTGCA	120
QY	136	GCGATATCTTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGATGTACTGGGAGATGA	195
Db	121	GGCCTATCTTTGAAGACCATGTCAGAAGAAGAAGCAAAATGAAGAGATGTACTGGGAGATGA	180
QY	196	AACAGAGGAAGAAGAACAAAGCCATTGAGCTCCCTGTCCTAAAGAGCAAGAACCCCTCTGA	255
Db	181	AACTGAGGAAGAAGAACCAAAGCCCATTAGAACTGCCCTGTTAAAGAGGAAGAACTCTCTGA	240
QY	256	AAAAACTGTTTGATGTGGCAGCAGAGAAGAAGTGGTGAATAATTACATCTGAAATACCACA	315
Db	241	AAAGGCTGTTTGATATGTCATCGGAAGAAAGAGGTGTAAGATTTACATCTGGGATACCTCA	300
QY	316	GACTGAGAGAAATGCAGAAGAGGGCTGAACGATTCAAATGTACCTGTGAGCTTGGAGAGTAA	375
Db	301	GACTGAAGAAGATGCAGAAGAGGGCTGAACGATTCAAATGTACCTGTAAGCCCTGGGAGAGTAA	360
QY	376	GAAGAGCTGCTCGGCACGTAGTTTGGATTTCTTTCAGTTTCCAAACAAAAGTCTGTCA TC	435
Db	361	GAAGGCTGCCCCGGCGCTAGTTTGGAAATTTCTTCAGTTTCCAAACAAAAGTTTATCATC	420
QY	436	TGATAACAAACCTATGTTTAACTTTGGATAAGCTGAAGGAAAGAGCTCAAAAGATTTGGTTT	495
Db	421	TGACACCAAGCCTATGCTTTAACTCTGGATAAATTAAGGAAAGAGCA CAGAGATTTGGTTT	480
QY	496	GAATGTCCTCTCAATCTCCGAAAGTCTGAAGATGTAGAAACTGAAAAGAGGAAGGA	555
Db	481	GAATGTCCTCTTCCATCTCTAGAAAGTCTGAGGATGATGAGAAGCTGAAGAAACCGGAAGGA	540
QY	556	GCGATTTGGGATTTGTCA CAAGTTTCAGCTGGAACTGGAAACCA CAGAGATATCAGAGGCAAA	615
Db	541	GAGATTTGGGATTTGTGACAAGTTTCAGCTGGAACTGGAAACCA CAGAGAGATATCAGAGGCAAA	600
QY	616	GAAGAGAAAGAGCAGAGCGCTTTGGATTTGCTGATGA AAAAGTTCTCTGATCTTTCTG	675
Db	601	GA AAAGAAAAGAGCAGAGCGTTTGGAAATGGCAATATGA AAGTCTCGTG--CTTTCTG	657

QY	676	TTCTCAGTGTTTCATTTCTCTCTCTCTGTCACATATATGCCTTAATGACACAG	735
Dd	658	CACCATAGTGTGTTTTCCAAATTTTCTCGTTGTTTATATATACATACACACCCCT	717
QY	736	TCATGTGGCTACGTC-----CTGCCCTCGCAATCAGGAGCATCTACCCCA	780
Dd	718	ACCTACCTACTATCTACATACACAGTCATGTGCTCCACAGTCAGAAGCATGTACGTTT	777
QY	781	GGTACATCCATGAACTCGGCAGCAGCTTTGACATTATGTCTGTTTTCAGCTTTAAGGTTGTT	840
Dd	778	GGTACATCTCTGNAATCAGTAGCAGGTTTATTAACTGCTCTCTCGGTTTAGATGTT	837
QY	841	GTGTTTTTGTGTTTGAATGTT 863	
Dd	838	GTGTTTTTGTGTTTAAATATT 860	

RESULT 15
AC022200/c
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-562017 map 8, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
AC022200 GI:13518207
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 148152)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-262017
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 148152)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P.,
DeAtreilano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galegan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marcuis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Menes,I., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliou,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

Direct Submission
Submitted (26-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2001 this sequence version replaced gi:13488035.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5196
Center clone name: 262_0_17
----- Summary Statistics
Sequencing vector: M13; M77815; 63% of reads
Sequencing vector: Plasmid; n/a; 37% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 146959 bases at least Q40
Consensus quality: 147487 bases at least Q30

Consensus quality: 147658 bases at least Q20
Insert size: 145000; agarose-fp
Quality size: 147852; sum-of-contigs
Quality coverage: 8.8 in Q20 bases; agarose-fp
Quality coverage: 8.7 in Q20 b.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

- * 1 40424: contig of 40424 bp in length
- * 40425 40524: gap of 100 bp
- * 40525 48708: contig of 8184 bp in length
- * 48709 48808: gap of 100 bp
- * 48809 73107: contig of 24299 bp in length
- * 73108 73207: gap of 100 bp
- * 73208 148152: contig of 74945 bp in length.

FEATURES

source
1. .148152
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-262017"
/clone_lib="RPC1-11 Human Male BAC"
1. .40424
/note="assembly_fragment"
clone_end:SP6
vector_side:left
40525. .48708
/note="assembly_fragment"
48809. .73107
/note="assembly_fragment"
73208. .148152
/note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN

Query Match 61.9%; Score 540.4; DB 2; Length 148152;
Best Local Similarity 86.9%; Pred. No. 1.3e-122;
Matches 643; Conservative 0; Mismatches 86; Indels 11; Gaps 4;
Qy 127 CAGACTCCAGCATATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAGATGTACT 186
Db 133780 CTGGCTCCAGGCATATTTTTCAGGGCATGCTGAAGAGG-----CAAATGAAGATGTACT 133727
Qy 187 GGGAGATGAACAGAGGAAGAAACAAAG-CCATTGAGCTCCCTGTCAAAGAGGAAG 245
Db 133726 GGGAGATGAACAGAGGAAGAAACAAAGCCCATAGAAATCCCTGTCAAAGAGGAAG 133667
Qy 246 AACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGGTCAAAATTACATCTG 305
Db 133666 AACTCCCTGAAAAAATCTGTTGACGTGGAAGCAGAGAGAAAGTGGTAAAAATTACATCTG 133607
Qy 306 AAATACACAGACTGAGAGAAATGAGAGAGGGCTGAACGATTCATATGTACCTGTGAGCT 365
Db 133606 AAATACACAGGAGAGAGAAATGAGAGAGGGCTGAACAGTCCATGTACCCGTGAGCT 133547
Qy 366 TGGAGGTGAAGAAAGCTCTGGCAGCTAGTGTGGGATTTCTTCAGTTCACAAAG 425
Db 133546 TGTAAGTGAAGAAAGCTCCCGGAGCTAGGTTGGGACTTCTTCAGTTTCAACAAAG 133487
Qy 426 GTCTGTATCTGATAACAACTATGTTAACTTGGATAGCTGAAGGAAAGAGCTCAAA 485
Db 133486 GTCTGTATCTAACAACAACTATCATGAACCTGGATAGCTAAGGAAAGAGCTCAA 133427
Qy 486 GATTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 545
|||||

Db 133426 GATTGGTTTGAATGGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 133367
Qy 546 AGAGGAAGAGCGGATTTGGGATTTGTCAAAAGTTCAGCTGGAACCTGGA---ACCACAGAG 602
133366 GGAGGAAGGAGCTGATTTGGGATTTATCACTGTTTCAGCTGGAACCCAGAAAGGAACACAGAGG 133307
Qy 603 ATACAGAGGCAAGAGAGGAGAAAGAGCAGAGCGCTTTGGGATTTGCCCTGATGAAAAGTTC 662
133306 ATACAGAGGCAAGAGATAGGAAAGAGCCTCGGCTATGGGATTTGCCCTGCTGAAAAGTTC 133247
Qy 663 CTGATACTTTCTGTTCTCCAGTGTTTTCCCATTTCTCTCCTTTCTTTGGTCCACATATATG 722
133246 ATGCTGCTTTCTGTTCTCCAGTGTGTTTCCATTTCTGTTGATTTCTTCTCGGTCACAATATA 133187
Qy 723 CCTAAATGCAAGTCAATGTCCTACGTCCTCGCCTCGCAATGAGGGAGCATGTACCCAGG 782
133186 CCTAAATGCAAGTCAAGTACCTACCTCTGCTTGAATGAGGAGGACATGTGTTCCCGAGC 133127
Qy 783 TACATCCATGACTGGCGCAGCAGTTTGACTTATGCTGTTTTCAGCTTTAAGGTTGTTGT 842
133126 TACATGTGGGAACCTCCAGCAGCAATTTGACTTATGATGTTCCAAC--TTAAGCTGTGT 133068
Qy 843 GTTTTGTGTTTGTATTATGT 862
Db 133067 GTTTTGTGTTTGTATTATTT 133048

Search completed: October 4, 2005, 22:06:49
Job time : 4052.85 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 22:07:00 ; Search time 730.217 Seconds
(without alignments)
8254.755 Million cell updates/sec

Title: US-09-788-476A-3
Perfect score: 873
Sequence: 1 TGGAGTgaggggtacaaga.....tgattatgttctgttaat 873

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7442561 seqs, 345328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	873	100.0	873	9	US-09-788-476A-3
2	873	100.0	894	9	US-09-788-476A-3
3	873	100.0	1154	9	US-09-764-846-90
4	873	100.0	1154	14	US-10-091-483-90K
5	853.4	97.8	891	9	US-09-822-830A-389
6	817.8	93.7	1400	21	US-10-956-157-7827
7	817.8	93.7	1818	21	US-10-956-157-2592

8	733.4	84.0	3147	16	US-10-006-285-419	Sequence 419, App
9	690.8	79.1	2553	10	US-09-814-353-20673	Sequence 20673, A
c 10	595.2	68.2	2553	10	US-09-814-353-20673	Sequence 20673, A
11	498.4	57.1	1022	16	US-10-006-285-233	Sequence 233, App
12	458.4	52.5	488	17	US-10-242-535A-29658	Sequence 29658, A
13	458.4	52.5	488	17	US-10-085-783A-29658	Sequence 29658, A
14	455	52.1	470	17	US-10-242-535A-28856	Sequence 28856, A
15	455	52.1	470	18	US-10-085-783A-28856	Sequence 28856, A
16	431	49.4	470	17	US-10-242-535A-40257	Sequence 40257, A
17	431	49.4	470	18	US-10-085-783A-40257	Sequence 40257, A
18	386	44.2	426	17	US-10-242-535A-40571	Sequence 40571, A
19	386	44.2	426	17	US-10-085-783A-40571	Sequence 40571, A
20	379	43.4	379	17	US-10-242-535A-43248	Sequence 43248, A
21	379	43.4	379	18	US-10-085-783A-43248	Sequence 43248, A
22	336.2	38.5	463	10	US-09-918-995-15429	Sequence 15429, A
23	334	38.3	408	9	US-09-930-213-64	Sequence 64, Appl
c 24	265.4	30.4	5469	9	US-09-764-877-4000	Sequence 4000, Ap
c 25	265.4	30.4	5469	10	US-09-764-891-9371	Sequence 9371, Ap
c 26	265.4	30.4	5469	15	US-10-205-428-817	Sequence 817, App
c 27	265.4	30.4	5469	17	US-10-242-515-4000	Sequence 4000, Ap
c 28	265.4	30.4	9453	9	US-09-764-877-3999	Sequence 3999, Ap
c 29	265.4	30.4	9453	10	US-09-764-891-9370	Sequence 9370, Ap
c 30	265.4	30.4	9453	15	US-10-205-428-816	Sequence 816, App
c 31	265.4	30.4	9453	17	US-10-242-515-3999	Sequence 3999, Ap
c 32	246.2	28.2	300	16	US-10-006-285-196	Sequence 196, App
33	244	27.9	275	20	US-10-425-115-158622	Sequence 158622, Sequence 531, App
34	211.8	24.3	255	10	US-09-930-213-531	Sequence 737, App
35	206	23.6	343	17	US-10-062-674-737	Sequence 1714, Ap
36	174	19.9	257	17	US-10-062-674-1714	Sequence 378, App
37	154.2	17.7	215	9	US-09-728-445-378	Sequence 378, App
38	154.2	17.7	215	22	US-10-964-549-378	Sequence 15759, A
39	92.2	10.6	463	10	US-09-814-353-15759	Sequence 26486, A
c 40	55	6.3	399	22	US-10-450-763-26486	Sequence 29210, A
41	55	6.3	399	22	US-10-450-763-29210	Sequence 32, Appl
42	54	6.2	5659	15	US-10-172-086-32	Sequence 348, App
43	54	6.2	5659	15	US-10-311-455-348	Sequence 40, Appl
44	54	6.2	5659	15	US-10-240-485-40	Sequence 14, Appl
45	54	6.2	5659	19	US-10-311-507-14	

ALIGNMENTS

RESULT 1
US-09-788-476A-3
; Sequence 3, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788,476A
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PCR Product
US-09-788-476A-3

Query Match	100.0%	Score 873;	DB 9;	Length 873;
Best Local Similarity	100.0%	Pred. No. 2.7e+232;		
Matches	873;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	TGGAGTgaggggtacaagaatgaggccgacgagcggtgagctccataagctaaagtgc	60	
Db	1	TGGAGTgaggggtacaagaatgaggccgacgagcggtgagctccataagctaaagtgc	60	
Qy	61	CGAACTAAGCAAGATGCTTCTGCTGTTGGAGACCAAGGAATAAGCAAGATCT	120	

Db 61 CGAACTAAAGCAAGATGTCTTGCTCGTGGTGGAGACCAAGGAAATAAGCAAGATCT 120
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGA 180
Db 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGA 180
Qy 181 TGTACTGGGAGATGAACAGAGGAAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 240
Db 181 TGTACTGGGAGATGAACAGAGGAAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 240
Qy 241 GGAAGAACCCCTCGAAATACTGTGTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTAC 300
Db 241 GGAAGAACCCCTCGAAATACTGTGTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTAC 300
Qy 301 ATCTGAAATACCAAGACTGAGAGAATGCAAGAGAGGGCTGAAAGATTCAATGTACTGTT 360
Db 301 ATCTGAAATACCAAGACTGAGAGAATGCAAGAGAGGGCTGAAAGATTCAATGTACTGTT 360
Qy 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCACAAC 420
Db 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCACAAC 420
Qy 421 AAAAGGTCTGCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGAGC 480
Db 421 AAAAGGTCTGCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGAGC 480
Qy 481 TCAAAGATTTGGTTGAAATGCTCTTCAATCTCGAGAAAGTCTGAAGATGATGAGAAACT 540
Db 481 TCAAAGATTTGGTTGAAATGCTCTTCAATCTCGAGAAAGTCTGAAGATGATGAGAAACT 540
Qy 541 GAAAAGAGCAAGAGAGCGATTTGGGATTTGACAAAGTTGAGCTGGAACCTGGAACCCACAGA 600
Db 541 GAAAAGAGCAAGAGAGCGATTTGGGATTTGACAAAGTTGAGCTGGAACCTGGAACCCACAGA 600
Qy 601 GGATACAGAGCAAGAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 660
Db 601 GGATACAGAGCAAGAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 660
Qy 661 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 661 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Qy 721 TGCTTAAATGCACAGTCATGCTAGCTCGCTCGCAATGAGGAGCATGACCCCA 780
Db 721 TGCTTAAATGCACAGTCATGCTAGCTCGCTCGCAATGAGGAGCATGACCCCA 780
Qy 781 GGTACATCCATGAACCTGCGGAGAGGAAAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 840
Db 781 GGTACATCCATGAACCTGCGGAGAGGAAAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 840
Qy 841 GTGTTTTTGTGTTTGAATATGTTGCTTGTAAAT 873
Db 841 GTGTTTTTGTGTTTGAATATGTTGCTTGTAAAT 873

RESULT 2

US-09-788-476A-1
; Sequence 1, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788, 476A
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 894
; TYPE: DNA
; ORGANISM: human
US-09-788-476A-1

Query Match 100.0%; Score 873; DB 9; Length 894;
Best Local Similarity 100.0%; Pred. No. 2,7e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGGAGTGAAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 60
Db 6 TGGAGTGAAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 65
Qy 61 CGAACTAAAGCAAGATGTCTTGCTCGTGGTGGAGACCAAGGGAATAAGCAAGATCT 120
Db 66 CGAACTAAAGCAAGATGTCTTGCTCGTGGTGGAGACCAAGGGAATAAGCAAGATCT 125
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGA 180
Db 126 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGA 185
Qy 181 TGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 240
Db 186 TGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 245
Qy 241 GGAAGAACCCCTCGAAATACTGTGTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTAC 300
Db 246 GGAAGAACCCCTCGAAATACTGTGTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTAC 305
Qy 301 ATCTGAAATACCAAGACTGAGAGAATGCAAGAGAGGGCTGAAAGATTCAATGTACTGTT 360
Db 306 ATCTGAAATACCAAGACTGAGAGAATGCAAGAGAGGGCTGAAAGATTCAATGTACTGTT 365
Qy 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCACAAC 420
Db 366 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCACAAC 425
Qy 421 AAAAGGTCTGCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGAGC 480
Db 426 AAAAGGTCTGCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGAGC 485
Qy 481 TCAAAGATTTGGTTGAAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACT 540
Db 486 TCAAAGATTTGGTTGAAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACT 545
Qy 541 GAAAAGAGCAAGAGAGCGATTTGGGATTTGCAAGTTGAGCTGGAACCTGGAACCCACAGA 600
Db 546 GAAAAGAGCAAGAGAGCGATTTGGGATTTGCAAGTTGAGCTGGAACCTGGAACCCACAGA 605
Qy 601 GGATACAGAGCAAGAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 660
Db 606 GGATACAGAGCAAGAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 665
Qy 661 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 666 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 725
Qy 721 TGCTTAAATGCACAGTCATGCTAGCTCGCTCGCAATGAGGAGCATGACCCCA 780
Db 726 TGCTTAAATGCACAGTCATGCTAGCTCGCTCGCAATGAGGAGCATGACCCCA 785
Qy 781 GGTACATCCATGAACCTGCGGAGAGGAAAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 840
Db 786 GGTACATCCATGAACCTGCGGAGAGGAAAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 845
Qy 841 GTGTTTTTGTGTTTGAATATGTTGCTTGTAAAT 873
Db 846 GTGTTTTTGTGTTTGAATATGTTGCTTGTAAAT 878

RESULT 3

US-09-764-846-90
; Sequence 90, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212

```
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1149)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-846-90

Query Match      100.0%; Score 873; DB 9; Length 1154;
Best Local Similarity 100.0%; Pred. No. 3.1e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGTGGGGTAAACAAGATGCGACCGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 60
DB 182 TGGAGTGGGGTAAACAAGATGCGACCGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 241
QY 61 CGAACTAAACCAAGATGCTTCTGCTGCTGTTGGAGACCAAGGAATAAGCAAGATCT 120
DB 242 CGAACTAAACCAAGATGCTTCTGCTGCTGTTGGAGACCAAGGAATAAGCAAGATCT 301
QY 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAATGAAGA 180
DB 302 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAATGAAGA 361
QY 181 TGACTGGGAGTAAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 420
DB 542 GAGCTTGGAGATGAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 601
QY 421 AAAAGGTCTGTCATCTGATAAACAACCTATGTTGGTAACTGGAAGAGGCTGGAACCT 480
DB 602 AAAAGGTCTGTCATCTGATAAACAACCTATGTTGGTAACTGGAAGAGGCTGGAACCT 721
QY 601 GATATACAGAGCAAGAAAGAGGAAAGAGAGGAGGCGTTTGGGATTCGCTGATGAAAGT 780
DB 841 GGATACAGAGCAAGAAAGAGGAAAGAGAGGAGGCGTTTGGGATTCGCTGATGAAAGT 841
QY 661 TCCTGATACATCTTCAGTGTTCCTGATTTTCCATTTCTCTCTCTCTCTCTCTCTCTCT 720
DB 842 TCCTGATACATCTTCAGTGTTCCTGATTTTCCATTTCTCTCTCTCTCTCTCTCTCTCT 901
QY 721 TGCTTAATGACAGTCAATGTCCTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
DB 902 TGCTTAATGACAGTCAATGTCCTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 961
QY 781 GGTACATCCATGACTCGGCGAGGAGTTGACTTATGCTTTCAGCTTTAAGTTGTT 840
```

```
Db 962 GGTACATCCATGAACTGGCGACAGCTTGACTTATGCTGTTTTCAGCTTTAAGTTGTT 1021
QY 841 GTCTTTTGTGTTTGTATGTTGCTTGTGTTTAAAT 873
DB 1022 GTCTTTTGTGTTTGTATGTTGCTTGTGTTTAAAT 1054

RESULT 4
US-10-091-483-90
; Sequence 90, Application US/10091483
; Publication No. US20030049650A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ12C1
; CURRENT APPLICATION NUMBER: US/10/091,483
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 348
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1149)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-091-483-90

Query Match      100.0%; Score 873; DB 14; Length 1154;
Best Local Similarity 100.0%; Pred. No. 3.1e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGTGGGGTAAACAAGATGCGACCGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 60
DB 182 TGGAGTGGGGTAAACAAGATGCGACCGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 241
QY 61 CGAACTAAACCAAGATGCTTCTGCTGCTGTTGGAGACCAAGGAATAAGCAAGATCT 120
DB 242 CGAACTAAACCAAGATGCTTCTGCTGCTGTTGGAGACCAAGGAATAAGCAAGATCT 301
QY 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAATGAAGA 180
DB 302 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAATGAAGA 361
QY 181 TGACTGGGAGTAAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 240
DB 362 TGACTGGGAGTAAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 421
QY 241 GGAAGACCCCTGAAAGAACTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 480
DB 422 GGAAGACCCCTGAAAGAACTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 481
QY 301 ATCTGAATACCAAGACTGAGAGATGCGAGAGAGGCTGGAACCTGGAACCTGGAACCT 360
DB 482 ATCTGAATACCAAGACTGAGAGATGCGAGAGAGGCTGGAACCTGGAACCTGGAACCT 541
QY 361 GAGCTTGGAGATGAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 420
DB 542 GAGCTTGGAGATGAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 601
QY 421 AAAAGGTCTGTCATCTGATAAACAACCTATGTTGGTAACTGGAAGAGGCTGGAACCT 480
DB 602 AAAAGGTCTGTCATCTGATAAACAACCTATGTTGGTAACTGGAAGAGGCTGGAACCT 661
QY 481 TCAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGCTGGAAGATGATGAAACT 540
DB 662 TCAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGCTGGAAGATGATGAAACT 721
QY 541 GAAAAGAGAGGAGGAGGAGGATTTGGGATTTGTCAAGTTTTCAGCTGGAACCTGGAAC 600
DB 722 GAAAAGAGAGGAGGAGGATTTGGGATTTGTCAAGTTTTCAGCTGGAACCTGGAAC 781
```

Qy	601	GGATACAGAGCGCAAGAGGAGGAAAAGAGCAGAGCGCTTTGGGATTTGGCTGATCAAAAGT	660
Db	782	GGATACAGAGCGCAAGAGGAGGAAAAGAGCAGAGCGCTTTGGGATTTGGCTGATCAAAAGT	841
Qy	661	TCTGTATCATTTTCGTGTTCTCCAGTGTGTTTCCATTCTCTCTCTTTGGTGACATATA	720
Db	842	TCTGTATCATTTTCGTGTTCTCCAGTGTGTTTCCATTCTCTCTCTTTGGTGACATATA	901
Qy	721	TGCTCTAAATGACAGTCATGTGCTCCTACGTCTCTGCTCGCAATAGGAGGAGCATGTACCCCA	780
Db	902	TGCTCTAAATGACAGTCATGTGCTCCTACGTCTCTGCTCGCAATAGGAGGAGCATGTACCCCA	961
Qy	781	GGTACATCCCATGAATCGGGCAGAGTTTGACTTATTGCTCTTTTCAGCTTTTAAGGTTGTT	840
Db	962	GGTACATCCCATGAATCGGGCAGAGTTTGACTTATTGCTCTTTTCAGCTTTTAAGGTTGTT	1021
Qy	841	GTGTTTTTGTGTTTTGATATATGTTCTGTTTAAT	873
Db	1022	GTGTTTTTGTGTTTTGATATATGTTCTGTTTAAT	1054

RESULT 5

```

US-09-822-830A-389/c
; Sequence 389, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 5402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 389
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-389

```

Query Match	97.8%	Score 853.4;	DB 9;	Length 891;
Best Local Similarity	99.8%	Pred. NO. 7.9e-227;		
Matches 865;	Conservative	0;	Mismatches 1;	Indels 1;
				Gaps 1;

Qy	8	AGGGTAAACAAGATGCGACCGACACGCTGCAGCTCCATAAGCTTAAAGCTTGC CGA ACTA	67
Db	886	AAGGGTTAACCAAGATGGCGACCGACACGCTGCAGCTCCATAAGCTTAAAGCTTGC CGA ACTA	827
Qy	68	AAGCAAGAAATGCTCTTGCTCGTGGTTTGAGACAAGGAAATAAAGCAAGATCTTTATCCAC	127
Db	826	AAGCAAGAAATGCTCTTGCTCGTGGTTTGAGACAAGGAAATAAAGCAAGATCTTTATCCAC	767

Qy	128	AGACTCAGGONTATCTTGAAGAACATGCTGAGAGGAGGCGAAATGAAGAAGATGTACTG	187
Db	766	AGACTCAGGONTATCTTGAAGAACATGCTGAGAGGAGGCGAAATGAAGAAGATGTACTG	707
Qy	188	GGAGATGAAACAGAGGAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAAGAGAGAA	247
Db	706	GGAGATGAAACAGAGGAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAAGAGAGAA	647
Qy	248	CCCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTACATCTGAA	307
Db	646	CCCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTACATCTGAA	587

308	ATACCA	CAGACTGAGAGAAATGCAGAGAGGGCTGAA	CGAATTCAATGATCACTGTGAGCTTG	367
586	ATACCA	CAGACTGAGAGAAATGCAGAGAGGGCTGAA	CGAATTCAATGATCACTGTGAGCTTG	527
368	GAGAGT	TAAGAAAGCTGCTCGGGCAGCTAGGTTTCGGATTTCTTCAGTTTCCAAACAAAGT		427
526	GAGAGT	TAAGAAAGCTGCTCGGGCAGCTAGGTTTCGGATTTCTTCAGTTTCCAAACAAAGT		467
428	CTGTCA	TCTCATTAACAAACCTATGTTTAACTTTGGATTAAGCTCGAAGGAAAGAGCTCAAAAGA		487
466	CTGTCA	TCTCATTAACAAACCTATGTTTAACTTTGGATTAAGCTCGAAGGAAAGAGCTCAAAAGA		407
488	TTTGGT	TTTGAATGTCCTCTTCAATCTCCAGAAAGTCTCGAAGATGATGAGAAACTGAAAAAG		547
406	TTTGGT	TTTGAATGTCCTCTTCAATCTCCAGAAAGTCTCGAAGATGATGAGAAACTGAAAAAG		347
548	AGGAAG	GAGCGATTTTGGGATTTGTACAAAGTTCACAGTTCAGCTGGAAC-TGGAAACACAGAGAGTAC		606
346	AGGAAG	GAGCGATTTTGGGATTTGTACAAAGTTCACAGTTCAGCTGGAAC-TGGAAACACAGAGAGTAC		287
607	AGAGGC	AAAGAGAGGAAAAAGACAGAGCGCTTTTGGGATTCGCTGATGAAAAAGTTCTCTGA		666
286	AGAGGC	AAAGAGAGGAAAAAGACAGAGCGCTTTTGGGATTCGCTGATGAAAAAGTTCTCTGA		227
667	TACTTT	CTGTTCTCCAGTGTTHCCATTTCTCTCTCTTCTTGGTTCACATATATGCGCTA		726
226	TACTTT	CTGTTCTCCAGTGTTHCCATTTCTCTCTCTTCTTGGTTCACATATATGCGCTA		167
727	AATGCA	CAGTCATGTGCTACGTCCTCGCTCGCAATGAGGAGAGCATGTACCCACGATACA		786
166	AATGCA	CAGTCATGTGCTACGTCCTCGCTCGCAATGAGGAGAGCATGTACCCACGATACA		107
787	TCCATGA	ACTCGGCGACAGTTTGACATTATGTGCTTTTTCAGCTTTTAAAGTGTGTGTGTTT		846
106	TCCATGA	CTCGGCGACAGTTTGACATTATGTGCTTTTTCAGCTTTTAAAGTGTGTGTGTTT		47
847	TTGTTTTT	TTGATTATGTTGCTGTTTAAT		873
46	TTGTTTTT	TTGATTATGTTGCTGTTTAAT		20

RESULT 6

```

US-10-956-157-7827
; Sequence 7827, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, Nucleic Acid Arrays F
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956, 157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7827
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-7827

```

	Query Match	93.7%	Score 817.8;	DB 21;	Length 1400;
	Best Local Similarity	99.8%;	Pred. No. 8.2e-217;		
	Matches 819;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	53	AAGCTTTCGCGAACTAAAGCAAGAATGCTTTCGCTGCTGTTTGGAGACCAAGGGAATAAAG	112		
Db	553	AAACTTTCGCGAACTAAAGCAAGAATGCTTTCGCTGCTGTTTGGAGACCAAGGGAATAAAG	612		
Qy	113	CAAGATCTTTATCCACAGACTCCAGGAGCATATCTTTGAAGAAATGCTGTAAGAGGAGGCAAAAT	172		
Db	613	CAAGATCTTTATCCACAGACTCCAGGAGCATATCTTTGAAGAAATGCTGTAAGAGGAGGCAAAAT	672		

[illegible]

RESULT 7
 US-10-956-157-2592
 ; Sequence 2592, Application US/10956157
 ; Publication No. US20050118625A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wyeth
 ; APPLICANT: Mounts, William
 ; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
 ; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
 ; FILE REFERENCE: 031896-043000 (AM 101081)
 ; CURRENT APPLICATION NUMBER: US/10/956.157
 ; CURRENT FILING DATE: 2004-10-04
 ; NUMBER OF SEQ ID NOS: 319805
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 2592
 ; LENGTH: 1818
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-956-157-2592

Query Match	93.7%	Score 817.8;	DB 21;	Length 1818;
Best Local Similarity	99.8%	Pred. No. 9.5e-217;		
Matches 819; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	53	AGCTTCCGCACTAAAGCAAGAATGCTTGTCTCGTGGTTTGGAGACCAAGGAAATAAG	112
Db	971	AAACTTCCGCAACTAAAGCAAGAATGCTTGTCTCGTGGTTTGGAGACCAAGGGAATAAG	1030
Qy	113	CAAGATCTTATCCACAGACTCCAGGCATATCTTTGAAGAAATGCTGAAGAGGAGGCAAAAT	172
Db	1031	CAAGATCTTATCCACAGACTCCAGGCATATCTTTGAAGAAATGCTGAAGAGGAGGCAAAAT	1090
Qy	173	GAGGAAGATGTACTGGGAGATGAAAACAGAGGAAGAGAAAACAAAGCCCATTTAGCTCCCT	232
Db	1091	GAGGAAGATGTACTGGGAGATGAAAACAGAGGAAGAGAAAACAAAGCCCATTTAGCTCCCT	1150
Qy	233	GTCAAGAGAGGAACCCCTCGAAAAAATCTTTGATGTGCAGCAGAGAGAAGAAGTGGTG	292
Db	1151	GTCAAGAGAGAGAACCCCTCGAAAAAATCTTTGATGTGCAGCAGAGAGAAGAAGTGGTG	1210
Qy	293	AAAATTAATCATCTGAAATPACACAGACTGAGAGAAATGCAGAAGAGGGCTGAACGATCAAT	352
Db	1211	AAAATTAATCATCTGAAATPACACAGACTGAGAGAAATGCAGAAGAGGGCTGAACGATCAAT	1270
Qy	353	GTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCA	412
Db	1271	GTACCTGTGAGCTTGGAGAGTAAGAAAGTGTCTCGGCAGCTAGGTTTGGGATTTCTTCA	1330
Qy	413	GTTCCTCAACAAAAGTCTGTCAATCTGTATAACAAACCTATGGTTAACTTGGATTAAGCTGAAG	472
Db	1331	GTTCCTCAACAAAAGTCTGTCAATCTGTATAACAAACCTATGGTTAACTTGGATTAAGCTGAAG	1390
Qy	473	GAAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGAT	532
Db	1391	GAAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGAT	1450
Qy	533	GAGAACTGAAAAGAGGAGGAGCGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGGA	592
Db	1451	GAGAACTGAAAAGAGGAGGAGCGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGGA	1510
Qy	593	ACCACAGAGGATACAGAGGCACAAAGAGGAAAAGAGCAGAGCGCTTTGGGATTCCTCGA	652
Db	1511	ACCACAGAGGATACAGAGGCACAAAGAGGAAAAGAGCAGAGCGCTTTGGGATTCCTCGA	1570
Qy	653	TGAAAAGTTCCTGATACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCTTCTTTGGT	712
Db	1571	TGAAAAGTTCCTGATACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCTTCTTTGGT	1630
Qy	713	CACATATATGCCCTAAATGCAAGTCATGTCCCTACGTCCTCGCAATGAGGGAGCAT	772
Db	1631	CACATATATGCCCTAAATGCAAGTCATGTCCCTACGTCCTCGCAATGAGGGAGCAT	1690
Qy	773	GTACCCAGGTTACATCCATGAACTGCGGCAGCAGTTTGACTTATTTGCTGTTTTCAGCTTTTA	832
Db	1691	GTACCCAGGTTACATCCATGAACTGCGGCAGCAGTTTGACTTATTTGCTGTTTTCAGCTTTTA	1750
Qy	833	AGGTTGTGTGTTTTTGTGTTTTTGATTAATGTTGTTTAAAT	873
Db	1751	AGGTTGTGTGTTTTTGTGTTTTTGATTAATGTTGTTTAAAT	1791

```

1  TITLE OF INVENTION: NOCULEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
2  TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
3  FILE REFERENCE: 031896-043000 (AM 101081)
4  CURRENT APPLICATION NUMBER: US/10/956.157
5  CURRENT FILING DATE: 2004-10-04
6  NUMBER OF SEQ ID NOS: 319805
7  SOFTWARE: PatentIn version 3.2
8  SEQ ID NO 2592
9  LENGTH: 1818
10 TYPE: DNA
11 ORGANISM: Homo sapiens
12 US-10-956-157-2592

```

Query Match	93.7%	Score 817.8;	DB 21;
Best Local Similarity	99.8%		
Pred. No. 9.5e-217;			
Matches 819;	Conservative	0;	Mismatches 2;
		Indels	0;
		Gaps	0;


```
Qy 753 GCCTGCAATCAGGAGCATGTACCCAGGTACA 786
Db 1375 GCCTGCAATCAGGAGCATGTACCCAGGTACA 1408

RESULT 10
US-09-814-353-20673/c
; Sequence 20673, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814.353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20673
; LENGTH: 2553
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-20673

Query Match 68.2%; Score 595.2; DB 10; Length 2553;
Best Local Similarity 96.9%; Pred. No. 1.1e-154;
Matches 628; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

Qy 202 GGAAGAGAAACAAGCCCATTTAGCTCCCTGTCAAAGAGGAAGAACCCCTG-AAAAA 260
Db 695 GGAAGAGAAACAAGCCCATTTAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 636

Qy 261 CTGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATACATCTGAATACCAAGACTG 320
Db 635 CTGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATACATCTGAATACCAAGACTG 576

Qy 321 AGAAGATGCAGAGAGGCTCAACGATTCATGTACCTGTGAGCTTGGAGAGTAAGAAAG 380
Db 575 AGAAGATGCAGAGAGGCTCAACGATTCATGTACCTGTGAGCTTGGAGAGTAAGAAAG 516

Qy 381 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCCAGTTCCAAACAAAAGGCTGTCTCATGATA 440
Db 515 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCCAGTTCCAAACAAAAGGCTGTCTCATGATA 456

Qy 441 ACAAACCTATCGTTAACTTTGGATTAAGCTGAAGGAAAGAGCTCAAGATTTGGTTGAATG 500
Db 455 ACAAACCTATCGTTAACTTTGGATTAAGCTGAAGGAAAGAGCTCAAGATTTGGTTGAATG 396

Qy 501 TCTCTCAATCTCAGAAAGCTCAGATGATGAGAACTGAAAGAGGAGGAGCGAT 560
Db 395 TCTCTCAATCTCAGAAAGCTCAGATGATGAGAACTGAAAGAGGAGGAGCGAT 336

Qy 561 TTGGATTTGTCAAGATTCAGCTGGAACCTGGAACCAACAGAGGATACAGAGCAAGAGA 620
Db 335 TTGGATTTGTCAAGATTCAGCTGGAACCTGGAACCAACAGAGGATACAGAGCAAGAGA 276

Qy 621 GGAAGAGCAGAGCGCTTTCGGATTCGCTGATGAAAGATTCCTGATACTTCTGTTCTC 680
Db 621 GGAAGAGCAGAGCGCTTTCGGATTCGCTGATGAAAGATTCCTGATACTTCTGTTCTC 680

Db 275 GGAAGAGCAGAGCGCTTTCGGATTCCTGATGAAAGATTCCTGATACTTCTGTTCTC 216
Qy 681 CAGTGTTCATATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 740
Db 215 CAGTGTTCATATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 156
Qy 741 TGCCTAGCTCTGCTCGCATGAGGAGCATGTACCCAGGTACATCCATGAAGTCCGG 800
Db 155 TGCCTAGCTCTGCTCGCATGAGGAGCATGTACCCAGGTACATCCATGAAGTCCGG 96
Qy 801 CAGCAGTTTGACTTATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 848
Db 95 GCACAGTTGACTTA-TGCTGTTTTCAGCTTAAAGTGTGTGTCGTTT 49

RESULT 11
US-10-006-285-233
; Sequence 233, Application US/10006285
; Publication No. US20030165854A1
; GENERAL INFORMATION:
; APPLICANT: Mary Jane Cunningham
; APPLICANT: Matthew R. Kaser
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS
; FILE REFERENCE: PA-0039 US
; CURRENT APPLICATION NUMBER: US/10/006,285
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PERL Program
; SEQ ID NO 233
; LENGTH: 1022
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165854A1 207354_Rn.3
; NAME/KEY: unsure
; LOCATION: 772-845
; OTHER INFORMATION: a, t, c, g, or other
US-10-006-285-233

Query Match 57.1%; Score 498.4; DB 16; Length 1022;
Best Local Similarity 87.3%; Pred. No. 6.4e-128;
Matches 580; Conservative 0; Mismatches 81; Indels 3; Gaps 3;

Qy 52 AAAGCTTTGCCCACTAAAGCAAGAAATGTTGCTGCTGGTTTGGAGACCAAGGAATAAA 111
Db 104 AAAGCTTTGCCCACTAAAGCAAGAAATGTTGCTGCTGGTTTGGAGACCAAGGAATAAA 163

Qy 112 GCAAGATCTTATCCACAGACT-CCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAA 170
Db 164 ACAAGATCTTATCAATAGGCTACCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAA 223

Qy 171 ATGAAGAGATGTACTGGGAGATGAACAGAGGAAGAAACAAACCCATTCAGCTCC 230
Db 224 ATGAAGAGATGTACTGGGAGATGAACAGAGGAAGAAACCAAGCCCTATGAAGTGC 283

Qy 231 CTGTCAAAGAGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGG 290
Db 284 CTGTCAAAGAGGAAGAACCCCTGAAAAAGTTGTTGATATGGCATCAGAAAGAGTGG 343

Qy 291 TGAAAAATTACATCTGAAATACCAAGAGAGATGAGAGAGAGGCTGAAAGATTC 350
Db 344 TAAAAATTACATCTGGAATACCTCAAACTGAGAGATGAGAGAGGCTGAAAGATTC 403

Qy 351 ATGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCG-GGCAGCTAGGTTTGGATTTCT 409
Db 404 ATGTGCTGTGAAGCTTGGAGAGTAAGAAAGGCTGCTCGCGCAGCAGGTTTGGAAATTTCT 463

Qy 410 TCAGTTCCAAACAAAGGCTGTCTCATCTCTGATAACAAACCTATGGTTAACTTGAAGCTG 469
Db 464 TCAGTTCCAAACAAAGGTTTATCATCTGACNCCAGCAATGGTTAACTTGAAGTAACTA 523

Qy 470 AAGGAAAGAGCTCAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGAT 529
```



```
Db 305 ACATCTGAATACACAGACTGAGAGAAATGCAGAAGAGGCTGAACGATTCAATGTACCT 364
Qy 359 GTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTTCCA 418
Db 365 GTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTTCCA 424
Qy 419 ACAAAGGTCTGTCATCTGATATAACAAACCTATGTTAACTTGGATTAAGCTGAAGGAAGA 478
Db 425 ACAAAGGTCTGTCATCTGATATAACAAACCTATGTTAACTTGGATTAAGCTGAAGGAAGA 484
Qy 479 GCTC 482
Db 485 GCTC 488

RESULT 14
US-10-242-535A-28856
; Sequence 28856, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28856
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (23)..(23)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-28856

Query Match 52.1%; Score 455; DB 17; Length 470;
Best Local Similarity 99.8%; Pred. No. 5.2e-116;
Matches 466; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TGGAGTGAGGGGTAAACAAG-ATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTG 59
Db 4 TGGAGTGAGGGGTAAACAAGNATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTG 63
Qy 60 CCGAACTAAAGCAAGAATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAGATC 119
Db 64 CCGAACTAAAGCAAGAATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAGATC 123
Qy 120 TTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAATGAAGAG 179
Db 124 TTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAATGAAGAG 183
Qy 180 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 239
Db 184 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 243
Qy 240 AGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTA 299
Db 124 TTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAATGAAGAG 183
Qy 180 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 239
Db 184 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 243
Qy 240 AGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTA 299
Db 244 AGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTA 303
Qy 300 CATCTGAAATACCAAGACTGAGAGAAATGCAGAAGAGGCTGAACGATTCAATGTACCTG 359
Db 184 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 243
Qy 240 AGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTA 299
Db 244 AGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTA 303
Qy 300 CATCTGAAATACCAAGACTGAGAGAAATGCAGAAGAGGCTGAACGATTCAATGTACCTG 359
```

```
Db 304 CATCTGAAATACCAAGACTGAGAGAAATGCAGAAGAGGCTGAACGATTCAATGTACCTG 363
Qy 360 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTTCCA 419
Db 364 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTTCCA 423
Qy 420 CAAAAGGTCTGTCATCTGATATAACAAACCTATGTTAACTTGGATTAAG 466
Db 424 CAAAAGGTCTGTCATCTGATATAACAAACCTATGTTAACTTGGATTAAG 470

RESULT 15
US-10-085-783A-28856
; Sequence 28856, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28856
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (23)..(23)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-28856

Query Match 52.1%; Score 455; DB 18; Length 470;
Best Local Similarity 99.8%; Pred. No. 5.2e-116;
Matches 466; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TGGAGTGAGGGGTAAACAAG-ATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTG 59
Db 4 TGGAGTGAGGGGTAAACAAGNATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTG 63
Qy 60 CCGAACTAAAGCAAGAATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAGATC 119
Db 64 CCGAACTAAAGCAAGAATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAGATC 123
Qy 120 TTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAATGAAGAG 179
Db 124 TTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAATGAAGAG 183
Qy 180 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 239
Db 184 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 243
Qy 240 AGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTA 299
Db 244 AGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAAAATTA 303
Qy 300 CATCTGAAATACCAAGACTGAGAGAAATGCAGAAGAGGCTGAACGATTCAATGTACCTG 359
Db 304 CATCTGAAATACCAAGACTGAGAGAAATGCAGAAGAGGCTGAACGATTCAATGTACCTG 363
Qy 360 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTTCCA 419
Db 364 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAAGTTCCA 423
Qy 420 CAAAAGGTCTGTCATCTGATATAACAAACCTATGTTAACTTGGATTAAG 466
```

Db 424 CAAAAGGCTGTGTCATCTGTGATAACAAACCTATGGTTAACTTGGATAAG 470

Search completed: October 5, 2005, 05:55:42
Job time : 735.217 secs


```
Db 6158 AATACCAAGGCTGAGAGAAATGAGAGAGAGGCGGCAAACTTCAGTGTACCTGTGAGCT 6099
Qy 366 TGGAGAGTAAGAAAGCTGCTGGGACAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAG 425
Db 6098 TGGAGAGTAAGAAAGCTGCTCAGGACAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAA - 6040
Qy 426 GTCTGTATCATGTATAACAAACCTATGTTTAACTTTGGATAGCTGGAAGGAAGAGCTCAAA 485
Db 6039 GTCTGTATCATGT--AACACACCTATGTTTAACTTTGGATAGCTGGAAGGAAGAGCTCAAA 5982
Qy 486 GATTGTGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 545
Db 5981 GATTGTGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 5922
Qy 546 AGAGGAAGAGCGGATTTGGGATTTGTCAAAAGTTTCAGCTGGAACTGGAAACACAGAGGATA 605
Db 5921 AGAGGAAGAGCGGATTTGGGATTTGTCAAAAGTTTCAGCTGGAACTGGAAACACAGAGGATA 5862
Qy 606 CAGAGGCAAAAGAGAGAAAGAGCAGAGCGCTTTGGGATGCTGATGAAAAGTTCCCTG 665
Db 5861 CAGAGGCAAAAGAGAGAAAGAGCAGAGCGCTTTGGGATGCTGATGAAAAGTTCCCTG 5802
Qy 666 ATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTTGGTCAATATATGCGCT 725
Db 5801 ATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTTGGTCAATATATACTA 5742
Qy 726 AATGCAAGTATGTCCTACGTCCTCGCAATGAGGAGAGCATGTACCCAGGTAC 785
Db 5741 AA-----TGCAGTCTTGCTTGCCTTGCATGAGGAGAGGTACCCAGGTAA 5698
Qy 786 ATCCATGAACCTGGGAGAGCTTGAATTTAGCTTATGCTGTTTCAAGCTTTAAGTTGTTGTT 845
Db 5697 AACTGTGAACCTGGGAGAGCTTGAATTTAGCTTATGCTTATGCTTTAAGTTGTTGTT 5638
Qy 846 TTTGTTTTTGATTATGTTGTTGTT 870
Db 5637 TTTGTTTTTGATTATGTTGTT 5613
```

RESULT 2

```
US-09-949-016-12964/c
; Sequence 12964, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12964
; LENGTH: 22303
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12964
```

```
Query Match 74.2%; Score 647.4; DB 4; Length 22303;
Best Local Similarity 90.4%; Pred No. 5.3e-189;
Matches 782; Conservative 0; Mismatches 46; Indels 37; Gaps 7;
Qy 7 GAGGGGTAAACAAGATGGCGGAGAGCGGTGGAGCTCCATAAGCTTAAAGCTTCCGAACT 66
Db 6441 GAGGGGTAAACAAGATGGTAACTGAGATGGTGGAGCTCCATAAGCTTGAAGCTTGTAACT 6382
```

```
Qy 67 AAAGCAAGAAATGCTTGTGCTGCTGTTGGAGACCAAGGGAATAAAAGCAAGATCTTATCCA 126
Db 6381 AAAGCAAGAAATGCTT-CTTGTGGTTTGGAGACCAAGGGAATAAAAGCAAGATCTTATCCA 6323
Qy 127 CAGACTCCAGGCAATATCTTGAAGAAACATGCTGAAGAGAGGCAAAATGAAGAAGATGTACT 186
Db 6322 CAGACTCCAGGCAATATCTTGAAGAAACATGCTGAAGAGAGGCAAAAT--GAAGATGTACT 6266
Qy 187 GGGAGATGAAACAGAGGAAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTGTCAAAGAGGAAGA 246
Db 6265 AGGAGATGAAACAGAGGAAGAAAGAA-----AACCCCTTGTCAAAATAGGAAGA 6219
Qy 247 ACCCCCTG-AAAGAACTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAATTTACATCTG 305
Db 6218 ACCCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAATTTACATCTG 6159
Qy 306 AATATCACACAGACTGAGAGAAATGCAAGAGAGGCTGAAACGATTTCAATGTACCTGTGAGCT 365
Db 6158 AATATCACACAGCTGAGAGAAATGCAAGAGAGGCGGCAACATTCAGTGTACCTGTGAGCT 6099
Qy 366 TGGAGATGAAAGAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAG 425
Db 6098 TGGAGATGAAAGAGCTGCTCAGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAA - 6040
Qy 426 GTCTGTATCATGTATCAACAACTATGTTTAACTTGGATAGCTGGAAGGAAGAGCTCAAA 485
Db 6039 GTCTGTATCATGT--AACACACCTATGTTTAACTTGGATAGCTGGAAGGAAGAGCTCAAA 5982
Qy 486 GATTGTGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 545
Db 5981 GATTGTGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 5922
Qy 546 AGAGGAAGAGCGATTTGGGATTTGTCAAAAGTTTCAGCTGGAACCTGGAACCAAGAGGATA 605
Db 5921 AGAGGAAGAGCGATTTGGGATTTGTCAAAAGTTTCAGCTGGAACCTGGAACCAAGAGGATA 5862
Qy 606 CAGAGCAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAGTTCCCTG 665
Db 5861 CAGAGCAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAGTTCCCTG 5802
Qy 666 ATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTTGGTCAATATATGCGCT 725
Db 5801 ATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTTGGTCAATATATACTA 5742
Qy 726 AATGCAAGTATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 785
Db 5741 AA-----TGCAGTCTTGCTTGCCTTGCATGAGGAGAGGTACCCAGGTAA 5698
Qy 786 ATCCATGAACCTGGGAGAGCTTGAATTTAGCTTATGCTGTTTCAAGCTTTAAGTTGTTGTT 845
Db 5697 AACTGTGAACCTGGGAGAGCTTGAATTTAGCTTATGCTTATGCTTTAAGTTGTTGTT 5638
Qy 846 TTTGTTTTTGATTATGTTGTTGTT 870
Db 5637 TTTGTTTTTGATTATGTTGTT 5613
```

RESULT 3

```
US-09-621-976-3627
; Sequence 3627, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3627
; LENGTH: 553
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..452
; US-09-621-976-3627

Query Match      60.1%; Score 525; DB 4; Length 553;
Best Local Similarity 99.8%; Pred. No. 4.1e-152;
Matches 536; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TGAGTGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 60
Db 17 TGAGTGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 76
Qy 61 CGAACTAAAGCAAGAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Db 77 CGAACTAAAGCAAGAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 136
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATCTGTAAGAGAGGAGGCAAAATGAAGAAGA 180
Db 137 TATCCACAGACTCCAGGCATATCTTGAAGAACATCTGTAAGAGAGGAGGCAAAATGAAGAAGA 196
Qy 181 TGTACTGGAGATGAACAAGAGGAAGAACAACAAGCCCATTTGAGCTCCCTGTGCAAGA 240
Db 197 TGTACTGGAGATGAACAAGAGGAAGAACAACAAGCCCATTTGAGCTCCCTGTGCAAGA 256
Qy 241 GGAAGAACCCCTGAAAAAAGTGTGATGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 257 GGAAGAACCCCTGAAAAAAGTGTGATGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 316
Qy 301 ATCTGAAATACCAAGACTGAGAGATGCAAGAGAGGCTGAAAGTTCATGTAATGTAACCTGT 360
Db 317 ATCTGAAATACCAAGACTGAGAGATGCAAGAGAGGCTGAAAGTTCATGTAATGTAACCTGT 376
Qy 361 GAGCTTGGAGATGAAGAAAGTGTCTGGGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 419
Db 377 GAGCTTGGAGATGAAGAAAGTGTCTGGGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 436
Qy 420 CAAAAGTGTGTATCTGATGAACAAAGCTATGGTTAACTTGGATTAAGCTGAAGGAAAGAG 479
Db 437 CAAAAGTGTGTATCTGATGAACAAAGCTATGGTTAACTTGGATTAAGCTGAAGGAAAGAG 496
Qy 480 CTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGCTGAAGATGATGAGA 536
Db 497 CTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGCTGAAGATGATGAGA 553

RESULT 4
US-09-513-999C-736
; Sequence 736, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 736
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..471
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: 354
; OTHER INFORMATION: y=c or t
US-09-513-999C-736

Query Match      53.1%; Score 465.6; DB 4; Length 471;
Best Local Similarity 99.8%; Pred. No. 9.2e-134;
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGAGTGGAGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 60
Db 6 TGAGTGGAGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 65
Qy 61 CGAACTAAAGCAAGAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Db 66 CGAACTAAAGCAAGAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATCTGTAAGAGAGGAGGCAAAATGAAGAAGA 180
Db 126 TATCCACAGACTCCAGGCATATCTTGAAGAACATCTGTAAGAGAGGAGGCAAAATGAAGAAGA 185
Qy 181 TGTACTGGGAGATGAACAAGAGGAAGAACAACAAGCCCATTTGAGCTCCCTGTGCAAGA 240
Db 186 TGTACTGGGAGATGAACAAGAGGAAGAACAACAAGCCCATTTGAGCTCCCTGTGCAAGA 245
Qy 241 GGAAGAACCCCTGAAAAAAGTGTGATGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 246 GGAAGAACCCCTGAAAAAAGTGTGATGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 305
Qy 301 ATCTGAAATACCAAGACTGAGAGATGCAAGAGAGGCTGAAAGTTCATGTAATGTAACCTGT 360
Db 306 ATCTGAAATACCAAGACTGAGAGATGCAAGAGAGGCTGAAAGTTCATGTAATGTAACCTGT 365
Qy 361 GAGCTTGGAGATGAAGAAAGTGTCTGGGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 366 GAGCTTGGAGATGAAGAAAGTGTCTGGGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 425
Qy 421 AAAAGGTCTGTCATCTGATTAACAAACCTATGTTAACTTGGATAAG 466
Db 426 AAAAGGTCTGTCATCTGATTAACAAACCTATGTTAACTTGGATAAG 471

RESULT 5
US-09-621-976-18639
; Sequence 18639, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18639
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 126..127
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-18639

Query Match      42.9%; Score 374.2; DB 4; Length 405;
Best Local Similarity 95.0%; Pred. No. 1.7e-105;
Matches 383; Conservative 11; Mismatches 7; Indels 2; Gaps 1;

Qy 2 GGAGTGGAGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 61
Db 1 GGAGTGGAGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 60
```

```
QY 62 GAACTAAAGCAAGATGTC--TTGCTGCTGTTTGGAGACCAAGGGAATAAAGCAAGATC 119
Db 61 GAACTAAAGCAAGATGTC--TTGCTGCTGTTTGGAGACCAAGGGAATAAAGCAAGATC 120
QY 120 TTATCCACAGACTCCAGGCATATCTTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAAG 179
Db 121 TKAIVNNACACTCCAGGCATATCTTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAAG 180
QY 180 ATGTACTGGAGATGAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 239
Db 181 ATGTACTGGAGATGAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 240 AGGAAGAACCCCTGAAAAAATCTTGTATGTCGACGAGAGAGAGAGAGAGAGAGAGAGAG 299
Db 241 AGGAAGAACCCCTGAAAAAATCTTGTATGTCGACGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 300 CATCTGAAATACACAGACTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
Db 301 CATCTGAAATACACAGACTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 360 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGGCAGCTAGTTTGG 402
Db 361 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGGCAGCTAGTTTGGG 403

RESULT 6
US-09-621-976-18638
; Sequence 18638, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S. J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18638
; LENGTH: 383
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18638
```

```
Query Match 38.0%; Score 331.6; DB 4; Length 383;
Best Local Similarity 99.7%; Pred. No. 2.5e-92;
Matches 331; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGTGAGGGGTAACAAAGATGGGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 60
Db 7 TGGAGTGAGGGGTAACAAAGATGGGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 66
QY 61 CGAACTAAAGCAAGATGTCCTGCTGCTGTTTGGAGCAAGGGAATAAAGCAAGATCT 120
Db 67 CGAACTAAAGCAAGATGTCCTGCTGCTGTTTGGAGCAAGGGAATAAAGCAAGATCT 126
QY 121 TATCCACAGACTCCAGGCATATCTTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAAG 180
Db 127 TATCCACAGACTCCAGGCATATCTTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAAG 186
QY 181 TGTAAGTGGAGATGAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 187 TGTAAGTGGAGATGAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
QY 241 GGAAGAACCCCTGAAAAAATCTTGTATGTCGACGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 247 GGAAGAACCCCTGAAAAAATCTTGTATGTCGACGAGAGAGAGAGAGAGAGAGAGAGAG 306
QY 301 ATCTGAAATACACAGACTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
Db 307 ATCTGAAATACACAGACTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 338
```

```
RESULT 7
US-09-621-976-18236
; Sequence 18236, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S. J.Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18236
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18236
```

```
Query Match 27.6%; Score 241; DB 4; Length 337;
Best Local Similarity 95.3%; Pred. No. 2.7e-64;
Matches 266; Conservative 3; Mismatches 8; Indels 2; Gaps 2;

QY 596 ACAGAGGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 655
Db 25 ACTTTGGAAACCACTAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 84
QY 656 AAAAGTTCCCTGATCTTCTGTTCTCCAGTGTTTCCAGTTCTCTCTCTCTCTCTCTCTCT 715
Db 85 AAAAGTTCCCTGATCTTCTGTTCTCCAGTGTTTCCAGTTCTCTCTCTCTCTCTCTCTCT 144
QY 716 ATATATGCTT-AAAATGCACAGTCATGTCCTAGCTGCTGCTGCTGCTGCTGCTGCTGCT 774
Db 145 ATATATGCTTAAAAAGCACAGTCATGTCCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 204
QY 775 ACCCAGGTACATCCATGAACCTGGCGCAGCAGTTTGACCTATGCTGCTGCTGCTGCTGCT 834
Db 205 ACCCAGGTACATCCATGAACCTGGCGCA-SAGTTTGACTATTGCTGCTGCTGCTGCTGCT 263
QY 835 GTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 873
Db 264 GTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 302
```

```
RESULT 8
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
```



```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgt-F1s
; US-08-232-463-14

Query Match          9.1%; Score 79.4; DB 1; Length 7218;
Best Local Similarity 6.3%; Pred. No. 2.1e-13;
Matches 26; Conservative 238; Mismatches 149; Indels 0; Gaps 0;

QY 2 GGAGTGAAGGTTAACAAGTGGCGACGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCC 61
DB 1433 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1374
QY 62 GAACTAAAGCAAGAATCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCTT 121
DB 1373 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1314
QY 122 ATCCACAGACTCCAGGATATCTTGAACAATCTGTAAGAGGCGCAATGAAGAAGAT 181
DB 1313 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1254
QY 182 GTACTGGAGATGAACAGAGGAGGAAGAACAAAGCCATTGAGCTCCCTGTCAAGAG 241
DB 1253 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1194
QY 242 GAAGAACCCCTGAAAAAAGTGTGATGTCGACAGAGAAAGTGGTCAAAATATCA 301
DB 1193 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1134
QY 302 TCTGAATACACAGACTGAGAGATGCAAGAGGCTGAACGATTAATGTAACCTGTG 361
DB 1133 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1074
QY 362 AGCTTGAGAGTAAGAAGCTGCTCGGCGAGCTAGGTTGGATTCTTCAGT 414
DB 1073 RRRRRRRATCGAAGCTCCCTCGACCTGCAGCCAAAGCTCGAATTAATCTGT 1021

RESULT 9
US-09-949-016-35088/c
; Sequence 35088, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42456
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-42456

Query Match          8.6%; Score 75; DB 4; Length 601;
Best Local Similarity 8.6%; Pred. No. 9.9e-13;
Matches 106; Conservative 0; Mismatches 10; Indels 5; Gaps 2;

QY 7 GAGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCCGA 66
DB 415 GAGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCTGA 356
QY 67 AAAGCAA---GAATGCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCTTA 122
DB 355 AAAGTAAGAATGAATGCTTCTGCTCGTGGTTGGAGA-CGAGCAAATAAGCAAGATTTT 297
QY 123 T 123
DB 296 T 296

RESULT 10
US-09-949-016-42456/c
; Sequence 42456, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42456
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-42456

Query Match          8.6%; Score 75; DB 4; Length 601;
Best Local Similarity 8.6%; Pred. No. 9.9e-13;
Matches 106; Conservative 0; Mismatches 10; Indels 5; Gaps 2;

QY 7 GAGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCCGA 66
DB 415 GAGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCTGA 356
QY 67 AAAGCAA---GAATGCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCTTA 122
DB 355 AAAGTAAGAATGAATGCTTCTGCTCGTGGTTGGAGA-CGAGCAAATAAGCAAGATTTT 297
QY 123 T 123
DB 296 T 296

RESULT 11
US-09-949-016-35089/c
; Sequence 35089, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgt-F1s
; US-08-232-463-14

Query Match          9.1%; Score 79.4; DB 1; Length 7218;
Best Local Similarity 6.3%; Pred. No. 2.1e-13;
Matches 26; Conservative 238; Mismatches 149; Indels 0; Gaps 0;

QY 2 GGAGTGAAGGTTAACAAGTGGCGACGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCC 61
DB 1433 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1374
QY 62 GAACTAAAGCAAGAATCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCTT 121
DB 1373 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1314
QY 122 ATCCACAGACTCCAGGATATCTTGAACAATCTGTAAGAGGCGCAATGAAGAAGAT 181
DB 1313 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1254
QY 182 GTACTGGAGATGAACAGAGGAGGAAGAACAAAGCCATTGAGCTCCCTGTCAAGAG 241
DB 1253 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1194
QY 242 GAAGAACCCCTGAAAAAAGTGTGATGTCGACAGAGAAAGTGGTCAAAATATCA 301
DB 1193 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1134
QY 302 TCTGAATACACAGACTGAGAGATGCAAGAGGCTGAACGATTAATGTAACCTGTG 361
DB 1133 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1074
QY 362 AGCTTGAGAGTAAGAAGCTGCTCGGCGAGCTAGGTTGGATTCTTCAGT 414
DB 1073 RRRRRRRATCGAAGCTCCCTCGACCTGCAGCCAAAGCTCGAATTAATCTGT 1021

RESULT 9
US-09-949-016-35088/c
; Sequence 35088, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42456
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-42456

Query Match          8.6%; Score 75; DB 4; Length 601;
Best Local Similarity 8.6%; Pred. No. 9.9e-13;
Matches 106; Conservative 0; Mismatches 10; Indels 5; Gaps 2;

QY 7 GAGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCCGA 66
DB 415 GAGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCTGA 356
QY 67 AAAGCAA---GAATGCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCTTA 122
DB 355 AAAGTAAGAATGAATGCTTCTGCTCGTGGTTGGAGA-CGAGCAAATAAGCAAGATTTT 297
QY 123 T 123
DB 296 T 296

RESULT 10
US-09-949-016-42456/c
; Sequence 42456, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```


Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:20:27 ; Search time 3485.58 Seconds
(without alignments)
9533.602 Million cell updates/sec

Title: US-09-788-476A-3

Perfect score: 873

Sequence: 1 tggagtggggtaacaaga.....tgattatgttgcttgaat 873

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848.2	97.2	871	5	BM916484
2	832.4	95.3	1000	7	CO580484
3	830.8	95.2	874	5	BM915580
4	825.2	94.5	877	5	BU849740
5	824.4	94.4	895	5	BU174287
6	824	94.4	1030	4	BM559381
7	820.4	94.0	963	7	CO648068
8	820	93.9	844	3	CF610192
9	820	93.9	910	3	AF161434
10	814.4	93.3	1070	4	BM460786
11	814	93.2	928	5	BU855435
12	811.4	92.9	817	5	BM456775
13	801.6	91.8	843	5	BM456776
14	798.8	91.5	823	5	BU599301
15	798.2	91.4	914	5	BU508603
16	797.4	91.3	910	5	BU157949
17	784.2	89.8	842	5	BQ692014
18	784	89.8	784	3	CR593273
19	783.8	89.8	937	5	BU856660
20	782	89.6	794	7	CN261593
21	782	89.6	1006	5	BQ068156
22	780.8	89.4	795	7	CN261585
23	780	89.3	806	4	BG574651
24	776.2	88.9	883	5	BU146608

25	768.6	88.0	779	5	BM916484
26	762.8	87.4	1105	7	BM916484
27	759	86.9	783	4	BK231375
28	758.8	86.9	1104	5	BM912983
29	758	86.8	914	5	BU902964
30	757	86.7	818	5	BQ227687
31	756.8	86.7	902	5	BQ622276
32	756.6	86.7	783	6	CA775512
33	755.8	86.6	760	4	BM722636
34	755	86.5	788	4	BM462456
35	753.6	86.3	876	5	BU855417
36	750.8	86.0	876	5	BU194990
37	749.4	85.8	783	4	BG533012
38	744.2	85.2	791	6	CB956735
39	742.2	85.0	878	5	BQ213428
c 40	741.2	84.9	785	5	BU632453
41	740.2	84.8	776	4	BI460375
42	738.2	84.6	887	4	BI255433
43	738	84.5	738	7	CF130101
44	738	84.5	819	5	BU596410
c 45	736.4	84.4	753	5	BM979030

ALIGNMENTS

RESULT 1
BM916484
LOCUS BM916484 871 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT 6641858 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5482902
5', mRNA sequence.
ACCESSION BM916484
VERSION BM916484.1 GI:19366863
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 871)
TITLE NIH-MGC http://mgs.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at:
http://image.llnl.gov
Plate: LLCM2009 row: 1 column: 07
High quality sequence stop: 748.

FEATURES

source
1. 871
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5482902"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 97.2% ; Score 848.2 ; DB 5 ; Length 871 ;

Best Local Similarity 99.5%; Pred. No. 4.7e-212;		Matches 850; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY	20	ATGGGACCGAGAGCGTGGAGCTCCATAAGCTTCCGAACTTAAGCAAGATGT	79
Db	1	ATGGGACCGAGAGCGTGGAGCTCCATAAGCTTAAGCTTCCGAACTTAAGCAAGATGT	60
QY	80	CTTGCTCGTGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCA	139
Db	61	CTTGCTCGTGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCA	120
QY	140	TATCTTTGAAGAACATGCTGAAGAGGCGCAATGAAGAGATGTACTGGAGATGAACA	199
Db	121	TATCTTTGAAGAACATGCTGAAGAGGCGCAATGAAGAGATGTACTGGAGATGAACA	180
QY	200	GAGAGAGAGAAACAAAGCCCATAGCTCCCTGTCMAAGAGAGAAACCCCTGAAAAA	259
Db	181	GAGAGAGAGAAACAAAGCCCATAGCTCCCTGTCMAAGAGAGAAACCCCTGAAAAA	240
QY	260	ACTGTTGATGTGCAGCAGAGAGAAAGTGCTGAAATTTACATCTGAAATACCACAGACT	319
Db	241	ACTGTTGATGTGCAGCAGAGAGAAAGTGCTGAAATTTACATCTGAAATACCACAGACT	300
QY	320	GAGAGATGCAGAGAGGCGTGAACGATTCATGTACCTGTGAGCTTTGGAGAGTAAGAA	379
Db	301	GAGAGATGCAGAGAGGCGTGAACGATTCATGTACCTGTGAGCTTTGGAGAGTAAGAA	360
QY	380	GCTGCTCGGCAGCTAGGTTGGATTTCTTCAGTTCCAAACAAAGTCTGTCTCAT	439
Db	361	GCTGCTCGGCAGCTAGGTTGGATTTCTTCAGTTCCAAACAAAGTCTGTCTCAT	420
QY	440	AACAAACCTATGTTAACTTGGATTAAGCTCAAGGAAAGACTCAAAAGATTTGGTTGAAT	499
Db	421	AACAAACCTATGTTAACTTGGATTAAGCTCAAGGAAAGACTCAAAAGATTTGGTTGAAT	480
QY	500	GTCTCTTCAATCTCCGAAAGTCTGAAGATGATGAGAACTGAGAAAGAGGAGGCGA	559
Db	481	GTCTCTTCAATCTCCGAAAGTCTGAAGATGATGAGAACTGAGAAAGAGGAGGCGA	540
QY	560	TTTGGGATTCACAGTTTCAGCTGGAACCTGGAACCAACAGAGATACAGAGCAAGAAG	619
Db	541	TTTGGGATTCACAGTTTCAGCTGGAACCTGGAACCAACAGAGATACAGAGCAAGAAG	600
QY	620	AGGAAAGACGACGCGCTTTGGATTTGCTGATGAAAGTTCTGATACCTTCTCTCT	679
Db	601	AGGAAAGACGACGCGCTTTGGATTTGCTGATGAAAGTTCTGATACCTTCTCTCT	660
QY	680	CCAGTGTCTTCCATTTCTCTCTCTTCTTCTGTCACATATATGCTAAATGCACAGTCAT	739
Db	661	CCAGTGTCTTCCATTTCTCTCTCTTCTTCTGTCACATATATGCTAAATGCACAGTCAT	720
QY	740	GTGCTACGTCCTGCGCAATGAGGAGCATGTATCCCGAGTACATCCATGAATGCG	799
Db	721	GTGCTACGTCCTGCGCAATGAGGAGCATGTATCCCGAGTACATCCATGAATGCG	780
QY	800	GCAGCAGTTTGACTTATGCTGTTTACGCTTAAAGTTGTTGTTTCTTTTCTGATTA	859
Db	781	GCAGCAGTTTGACTTATGCTGTTTACGCTTAAAGTTGTTGTTTCTTTTCTGATTA	840
QY	860	TGTTGCTTGTGTAAT	873
Db	841	TGGTGCTTGTGTAAT	854
RESULT 2			
CO580484			
LOCUS			
DEFINITION			
ILLUMIGEN MCQ 48614 Katze MMTF Macaca mulatta cDNA clone			
TBIUW:18168 57 similar to 'Bases 522 to 900 highly similar to human			
C1P29 (Hb.410597), mRNA sequence.			
ACCESSION			
CO580484			
VERSION			
CO580484.1 GI:50411638			
KEYWORDS			
EST.			
EST 20-JUL-2004			
1000 bp mRNA linear			
SOURCE			
Macaca mulatta (rhesus monkey)			
ORGANISM			
Macaca mulatta			
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;			
Cercopitheciinae; Macaca.			
REFERENCE			
1 (bases 1 to 1000)			
AUTHORS			
Katze, M.G., Thomas, M., Korth, M., Iadonato, S.P. and Magness, C.L.			
TITLE			
Large-scale Rhesus Macaque cDNA Sequencing			
JOURNAL			
Unpublished (2003)			
COMMENT			
Contact: C. Magness			
Illumigen Biosciences Inc.			
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA			
Tel: 2063780400			
Fax: 2063780408			
Email: cmagness@illumigen.com			
Sequenced on 2004.07.02. 775 Q20 bases. Library Preparation: Prof.			
Michael Katze Lab at University of Washington DNA Sequencing:			
Illumigen Biosciences Inc. For further information, see			
http://www.macaque.org			
PCR Primers			
FORWARD: CCTCACTAAAGGGACAAAA			
BACKWARD: CACTATAGGGCAATTTGGTA			
Insert length: 1000 Std Error: 0.00			
Plate: CL000408 row: B column: 10			
Seq primer: CCTCACTAAAGGGACAAAA			
POLYA=Yes.			
Location/Qualifiers			
1..1000			
/organism="Macaca mulatta"			
/mol_type="mRNA"			
/strain="Indian"			
/db_xref="taxon:9544"			
/clone="TBIUW:18168"			
/sex="male"			
/tissue_type="testis"			
/dev_stage="juvenile"			
/lab_host="Electromax DH10B"			
/clone_lib="Katze MMTF"			
/note="Vector: pDONR 222; Site 1: BsrG I; Site 2: BsrG I;			
Created from Cloneminer cDNA Library Construction kit			
(catalog #18249-029)"			
ORIGIN			
Query Match 95.3%; Score 832.4; DB 7; Length 1000;			
Best Local Similarity 98.1%; Pred. No. 7e-208;			
Matches 842; Conservative 0; Mismatches 16; Indels 0; Gaps 0;			
QY	16	CAAGATGGCGACGAGACGCTGAGCTCCATAAGCTTAAAGCTTCCGAACTAAAGCAAGA	75
Db	5	CAAGATGGCGACGAGACGCTGAGCTCCATAAGCTTAAAGCTTAAAGCAAGA	64
QY	76	ATGCTCTTGTCTGTTGGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCA	135
Db	65	ATGCTCTTGTCTGTTGGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCA	124
QY	136	GCATATCTTTGAAGAACATGCTGAAGAGGAGCAATGAAGAGATGTACTGGAGATGA	195
Db	125	GCATATCTTTGAAGAACATGCTGAAGAGGAGCAATGAAGAGATGTACTGGAGATGA	184
QY	196	AACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGA	255
Db	185	AACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGA	244
QY	256	AAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTGAAATACACA	315
Db	245	AAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTGAAATACACA	304
QY	316	GACTGAGAGAAATGCAGAGAGGCGTCAACGATTCAATGTACCTGTGAGCTTTGAGAGTAA	375
Db	305	GACTGAGAGAAATGCAGAGAGGCGTCAACGATTCAATGTACCTGTGAGCTTTGAGAGTAA	364
QY	376	GAAAGCTGCTCGGCGCAGCTAGGTTTCGGATTCTTCAGTTCCAAAGGTTCTGTCATC	435

RESULT 4

BUS49740 877 bp mRNA linear EST 16-OCT-2002
LOCUS AGENCOURT 10440969 NIH_MGC_109 Homo sapiens cDNA clone
DEFINITION IMAGE:6598412 5', mRNA sequence.

ACCESSION BUS49740

VERSION BUS49740.1 GI:24034703

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 877)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM2831 row: c column: 20

High quality sequence stop: 659.

FEATURES

source

1..877

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6598412"

/tissue_type="teratocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_109"

/notes="Organ: ovary; Vector: pOT7; Site 1: EcoRI; Site 2:

XhoI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCACAG(G). Library constructed by Ling Hong in the

Laboratory of Gerald M. Rubin (University of California,

Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and

Superscript II RT (Life Technologies). Note: this is a

NIH_MGC Library."

ORIGIN

Query Match 94.5%; Score 825.2; DB 5; Length 877;
Best Local Similarity 98.1%; Pred. No. 5.3e-206;
Matches 833; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 23 GCGACCGAGCGTGGAGCTCCATAGCTAAAGCTTCCGAACTAAAGCAAGATCTCTT 82

DB 1 GCGACCGAGCGTGGAGCTCCATAGCTAAAGCTTCCGAACTAAAGCAAGATCTCTT 60

QY 83 GCTCTGTTTGGAGACCAAGGCAATTAAGCAAGATCTTATCCACAGACTCCAGGCATAT 142

DB 61 GCTCTGTTTGGAGACCAAGGCAATTAAGCAAGATCTTATCCACAGACTCCAGGCATAT 120

QY 143 CTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGATGATCTGGAGATGAACAGAG 202

DB 121 CTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGATGATCTGGAGATGAACAGAG 180

QY 203 GAAGAGAACCAAGCCCATTTAGCTCCCTGTCAAGAGAGAAACCCCTGAAGAACT 262

DB 181 GAAGAGAACCAAGCCCATTTAGCTCCCTGTCAAGAGAGAAACCCCTGAAGAACT 240

QY 263 GTTGTATGTGCGACGAGAGAAAGTGGTGAATTTACATCTCAAAATACCACAGACTGAG 322

DB 241 GTTGTATGTGCGACGAGAGAAAGTGGTGAATTTACATCTCAAAATACCACAGACTGAG 300

QY 323 AGAATGCAAGAGGGCTGAACGATTCATGTCCTGTGAGCTTGAGAGTAAAGAGCT 382

DB 301 AGAATGCAAGAGGGCTGAACGATTCATGTCCTGTGAGCTTGAGAGTAAAGAGCT 360

QY 383 GCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAGGCTCTGTCACTGATAAC 442

DB 361 GCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAGGCTCTGTCACTGATAAC 420

QY 443 AAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAGCTCAAAGATTTGGTTTGAATGTC 502

DB 421 AAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAGCTCAAAGATTTGGTTTGAATGTC 480

QY 503 TCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGGAGGAGCGATTT 562

DB 481 TCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGGAGGAGCGATTT 540

QY 563 GGGATTGTGCACAAAGTTTCAGCTGGAACCTGGAACACACAGAGGATACACAGGCAAGAGAGG 622

DB 541 GGGATTGTGCACAAAGTTTCAGCTGGAACCTGGAACACACAGAGGATACACAGGCAAGAGAGG 600

QY 623 AAAAGAGCAGAGCGCTTTGGGATTCCTGATGAAAGTTTCTGATATCTTTCTGTTCTCCA 682

DB 601 AAAAGAGCAGAGCGCTTTGGGATTCCTGATGAAAGTTTCTGATATCTTTCTGTTCTCCA 660

QY 683 GTGTTTTCATTTCTCTCTTCTTGTGTCATATATGCTTAATGACAGTCAATGTC 742

DB 661 GTGTTTTCATTTCTCTCTTCTTGTGTCATATATGCTTAATGACAGTCAATGTC 720

QY 743 CTTAGCTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACCTGCGGCA 802

DB 721 CTTAGCTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACCTGCGGCA 780

QY 803 GCAGTTTGACTTATTGCTGTTTTCAGCTTTAAAGTTTGTGTTTGTGTTTGTGATTATGT 862

DB 781 GCAGTTTGACTTATTGCTGTTTTCAGCTTTAAAGTTTGTGTTTGTGTTTGTGATTATGT 840

QY 863 TCGTTGTGA 871

DB 841 GCTTGTGNTA 849

RESULT 5

BUI74287

LOCUS

DEFINITION

AGENCOURT 7994057 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6082078

5', mRNA sequence.

ACCESSION BUI74287

VERSION BUI74287.1 GI:22688271

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 895)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM2310 row: a column: 23

High quality sequence stop: 650.

Location/Qualifiers

1..895

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6082078"

/tissue_type="ductal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

FEATURES

source

[illegible]


```

Qy 729 TGCAAGTCAATGCTCTACGCTCTGCTCCCAATGAGGGAGCATGTAACCCAGGTACATC 788
Db 723 TGCAAGTCAATGCTCTACGCTCTGCTCCCAATGAGGGAGCATGTAACCCAGGTACATC 782
Qy 789 CATGAAGTGGGAGGAGTGTGACTTATCTGTTTTCAGCTTTAAGGTTGCTGTTT 848
Db 783 CGTGAAGTGGGAGGAGTGTGACTTATCTGTTTTCAGCTTTAAGGTTGTTGTTT 842
Qy 849 GTTTT 854
Db 843 TGTTT 848

CR610192 844 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CSOCAP004Yall of Thymus of Homo sapiens
(human).
CR610192
ACCESSION CR610192.1 GI:50490999
VERSION HTC; CNSLT cDNA.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 844)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
TITLE Full-length cDNA libraries and normalization
JOURNAL
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 844)
Genoscope.
Direct Submission
TITLE Direct Submission
JOURNAL
SUBMITTED (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
LOCATION/Qualifiers
FEATURES
source
1. .844
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOCAP004Yall"
/tissue_type="Thymus"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 93.9%; Score 820; DB 3; Length 844;
Best Local Similarity 100.0%; Pred. No. 1.2e-204;
Matches 820; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 AGCTTGCAGAACTAAAGCAAGATGCTTGTCTGCTGTTTGGAGACCAAGGGAATAAGC 113
Db 1 AGCTTGCAGAACTAAAGCAAGATGCTTGTCTGCTGTTTGGAGACCAAGGGAATAAGC 60

Qy 114 AAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAATGCTGTAAGAGAGGCAATG 173
Db 61 AAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAATGCTGTAAGAGAGGCAATG 120

Qy 174 AAGAGATGTTACTGGAGATGAACAGAGGAGAGAAACAAAGCCATTGAGTCCCTG 233
Db 121 AAGAGATGTTACTGGAGATGAACAGAGGAGAGAAACAAAGCCATTGAGTCCCTG 180

Qy 234 TCAAGAGGAGAACCCCTGAAACAACTGTTGATGTCGACGAGAGAAAGTGGTGA 293
Db 181 TCAAGAGGAGAAACCCCTGAAACAACTGTTGATGTCGACGAGAGAAAGTGGTGA 240

```

```

Qy 294 AAATTACATCTGAATATACACAGACTGAGAGATGAGAGAGGGCTGACGATTCAATG 353
Db 241 AAATTACATCTGAATATACACAGACTGAGAGATGAGAGAGGGCTGACGATTCAATG 300
Qy 354 TACCTGTGAGCTTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAG 413
Db 301 TACCTGTGAGCTTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAG 360
Qy 414 TTCCAAACAAAGGCTCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGGATAAGCTGAAG 473
Db 361 TTCCAAACAAAGGCTCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGGATAAGCTGAAG 420
Qy 474 AAAGAGCTCAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGCTGGAAGATGATG 533
Db 421 AAAGAGCTCAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGCTGGAAGATGATG 480
Qy 534 AGAACTGAAAGAGAGAGGAGGAGGATTTGGGATTTGTCAACAGTTTCAGCTGGAAGCTGAA 593
Db 481 AGAACTGAAAGAGAGAGGAGGAGGATTTGGGATTTGTCAACAGTTTCAGCTGGAAGCTGAA 540
Qy 594 CCACAGAGATACAGAGGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 653
Db 541 CCACAGAGATACAGAGGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
Qy 654 GAAAAGTTCCTGATATCTTCTGTTTCCAGTGTTCCTTCCATTTCTCTCTCTCTCTCTCTCT 713
Db 601 GAAAAGTTCCTGATATCTTCTGTTTCCAGTGTTCCTTCCATTTCTCTCTCTCTCTCTCTCT 660
Qy 714 ACATATATGCTTAATGACAGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 773
Db 661 ACATATATGCTTAATGACAGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Qy 774 TACCCAGGTACATCCATGAATGCGGCGACAGTTTGACTTATGCTGTTTTCAGCTTTAA 833
Db 721 TACCCAGGTACATCCATGAATGCGGCGACAGTTTGACTTATGCTGTTTTCAGCTTTAA 780
Qy 834 GGTTGTTGCTTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTT 873
Db 781 GGTTGTTGCTTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTT 820

RESULT 9
AF161434 910 bp mRNA linear HTC 22-MAY-2001
LOCUS Homo sapiens HSPC316 mRNA, partial cds.
DEFINITION AF161434
ACCESSION AF161434.1 GI:6841281
VERSION HTC.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 910)
Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
Human partial CDS from cd34+ stem cells
Unpublished
TITLE 2 (bases 1 to 910)
JOURNAL
REFERENCE Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
AUTHORS Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
Direct Submission
TITLE Submitted (14-MAY-1999) Shanghai Institute of Hematology, Shanghai
JOURNAL Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,
Shanghai 200025, P. R. China
FEATURES
source
1. .910
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CBLAIH07"
/cell_type="cd34+ stem cell"
/tissue_type="blood"
<1. .765

```

CDS

/codon_start=2	
/product="HSPC316"	
/protein_id="AAP28994.1"	
/db_xref="GI:6841282"	
/translation="SGVRGNKMATETVELHLKLAEKOECLARGLETGKIQKDLIHR	
LQAYLEHAEEANEEDVLGDETEEBETKPIELPVKEEPPPERTVDAEAKRVKXITS	
EIPTQRMQRARFNVFVLSLEKKAARAVGISVVPNKRSLITLNLWLWISLKE	
ELKDLGLNVSSIIRKSEDDKLLKRRKRFVITSSAGTGTTEARRRGEQSALGLP	
DEKFLILSLVLCQFPFLSFLLVTPMPKCTVNCILPASQ"	
ORIGIN	
Query Match	93.9%; Score 820; DB 3; Length 910;
Best Local Similarity	99.0%; Pred. No. 1.3e-204;
Matches 867; Conservative	0; Mismatches 5; Indels 4; Gaps 4;
QY	1 TGGAGTGAGGGTTAAACAGATGCGGACCGAGAGCGGTGGAGCTCCATAAGCTTAAGCTTGC 60
DB	4 TGGAGTGAGGGTTAAACAGATGCGGACCGAGAGCGGTGGAGCTCCATAAGCTTAAGCTTGC 63
QY	61 CGAACTAAAGCAAGATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAGATCT 120
DB	64 CGAACTAAAGCAAGATGTCTTGTCTGTGTTTGGAGACCAAGGGAATAAGCAAGATCT 123
QY	121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAAAATGAAGA 180
DB	124 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAAAATGAAGA 183
QY	181 TGTACTGGGAGATGAACAGAGAGAGAGAAACAAAGCCCAATTTGAGCTCCCTGTCAA 240
DB	184 TGTACTGGGAGATGAACAGAGAGAGAGAAACAAAGCCCAATTTGAGCTCCCTGTCAA 243
QY	241 GGAAGAACCCCTGAAAAAAGTGTGATGTGGCAGCAGAGAAAGTGGTGAATAATTAC 300
DB	244 GGAAGAACCCCTGAAAAAAGTGTGATGTGGCAGCAGAGAAAGTGGTGAATAATTAC 303
QY	301 ATCTGAAATACACAGACTGAGAGATGCGAAGAGGGCTGGAACGATTTCAATGTACTGT 360
DB	304 ATCTGAAATACACAGACTGAGAGATGCGAAGAGGGCTGGAACGATTTCAATGTACTGT 363
QY	361 GAGCTGGAGAGTAAGAAAGCTGCTCGGAGCAGTAGGTTGGGATTTCTTCAGTT-CCAA 419
DB	364 GAGCTGGAGAGTAAGAAAGCTGCTCGGAGCAGTAGGTTGGGATTTCTTCAGTTCCCAA 423
QY	420 CAAAAGTCTGTGAT-CTGATAACAAACCTATGTTAACTTGAATGAGCTGAAGGAAGA 478
DB	424 CAAAAGTCTGTGATCTGATTAACAAACCTATGTTAACTTGAATGAGCTTGAAGGAAGA 483
QY	479 GCTCAAGAATTT-GGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAA 537
DB	484 GCTCAAGAATTTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAA 543
QY	538 ACTGAAAAGAGGAGAGGAGATTTGGGATTTGTCACAAAGTTGAGCTGGAACCTGAA 597
DB	544 ACTGAAAAGAGGAGAGGAGATTTGGGATTTGTCACAAAGTTGAGCTGGAACCTGAA 603
QY	598 AGAGGATACAGAGCAAGAGAGAGAAAGAGCAGAGCGCTTTGGGATTCCTGATGAAA 657
DB	604 AGAGGATACAGAGC-AAGAGAGAGAAAGAGCAGAGCGCTTTGGGATTCCTGATGAAA 662
QY	658 AGTTCTGTGATCTTCTGTTCTCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCT 717
DB	663 AGTTCTGTGATCTTCTGTTCTCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCT 722
QY	718 ATATGCTTAAATGACAGTCAATGTGCTACGTCCTGCTGCAATGAGGGAGCATGTACC 777
DB	723 ATATGCTTAAATGACAGTCAATGTGCTACGTCCTGCTGCAATGAGGGAGCATGTACC 782
QY	778 CCAGGTACATCCATGACTCGGCAGCAGTTGACTTATTGCTGTTTTCAGCTTTAAGGTT 837
DB	783 CCAGGTACATCCATGACTCGGCAGCAGTTGACTTATTGCTGTTTTCAGCTTTAAGGTT 842
QY	838 GTTGTGTTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 873

Db	843 GTTGTGTTTTTGTGTTTTTGTGATTATGTTGCTTGTGTAAT 878
RESULT 10	
BM460786	
LOCUS	
DEFINITION	
AGENCOURT_6421520 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532141	
5', mRNA sequence.	
ACCESSION	
BM460786	
VERSION	
BM460786.1 GI:18509826	
KEYWORDS	
EST.	
SOURCE	
ORGANISM	
Homo sapiens (human)	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	
1 (bases 1 to 1070)	
NIH-MGC http://mgc.nci.nih.gov/.	
National Institutes of Health, Mammalian Gene Collection (MGC)	
Unpublished (1999)	
AUTHORS	
Contact: Robert Strausberg, Ph.D.	
TITLE	
Unpublished (1999)	
JOURNAL	
COMMENT	
Email: cgabbs-r@mail.nih.gov	
Tissue Procurement: ARCC	
cDNA Library Preparation: Life Technologies, Inc.	
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)	
Cloned by: Agencourt Bioscience Corporation	
Cloned distribution: MGC clone distribution information can be	
found through the I.M.A.G.E. Consortium/LLNL at:	
http://image.llnl.gov	
Plate: LLAM12214 row: o column: 22	
High quality sequence stop: 593.	
FEATURES	
Location/Qualifiers	
1..1070	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="IMAGE:5532141"	
/issue_type="leiomyosarcoma"	
/lab_host="DH10B (phage-resistant)"	
/clone_lib="NIH_MGC_71"	
/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2.1 kb."	
ORIGIN	
Query Match	93.3%; Score 814.4; DB 4; Length 1070;
Best Local Similarity	98.7%; Pred. No. 3.9e-203;
Matches 852; Conservative	0; Mismatches 7; Indels 4; Gaps 3;
QY	10 GGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGGCCGAATAAA 69
DB	4 GGGGAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGGCCGAATAAA 63
QY	70 GCAAGAATCTCTGCTGCTGGTTTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAG 129
DB	64 GCAAGAATCTCTGCTGCTGGTTTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAG 123
QY	130 ACTCCAGGATATCTTGAAGACATGCTCAAGAGGAGGCAATGAAGAGATGATCTGGG 189
DB	124 ACTCCAGGATATCTTGAAGACATGCTCAAGAGGAGGCAATGAAGAGATGATCTGGG 183
QY	190 AGATGAACAGAGGAGAAAGAAACAAAGCCCTTGGCTCCCTGTCAAAGAGGAAGAAC 249
DB	184 AGATGAACAGAGGAGAAAGAAACAAAGCCCTTGGCTCCCTGTCAAAGAGGAAGAAC 243
QY	250 CCTGAAAAAAGCTGTTGATGTGGCAGACAGAGAAAGTGGTGAATAATTAATCTGAAAT 309
DB	244 CCTGAAAAAAGCTGTTGATGTGGCAGACAGAGAAAGTGGTGAATAATTAATCTGAAAT 303
QY	310 ACCACAGACTGAGAGATGACAGAGAGGCTGACCGATTCAATGTTACCTGTGAGCTTGA 369
DB	304 ACCACAGACTGAGAGATGACAGAGAGGCTGACCGATTCAATGTTACCTGTGAGCTTGA 363
QY	370 GAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAAAGAGGTCT 429


```
RESULT 12
BX456775/c
LOCUS BX456775 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004YA11 linear EST 06-MAY-2004
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX456775
VERSION BX456775.2 GI:47071640
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 817)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 22, 2003 this sequence version replaced gi:31034801.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 6407.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0CAP004AA06NP1&c=6407.r.
FEATURES
source
1..817
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP004YA11"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
ORIGIN
Query Match 92.9%; Score 811.4; DB 5; Length 817;
Best Local Similarity 99.1%; Pred. No. 2.2e-202;
Matches 810; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 54 AGCTTGCAGCACTAAAGCAAGATGCTTCTGCTGCTGTTGGAGACCAAGGGAATAAGC 113
DB 817 AGCTTGCAGCACTAAAGCAAGATTTCTTBGCTCGTGGTTTGGAGACCAAGGGAATAAGC 758
QY 114 AAGATCTTATCCAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAAAATG 173
DB 757 AAGATCTTATCCAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAAAATG 698
QY 174 AAGAAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTAGCTCCCTG 233
DB 697 AAGAAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTAGCTCCCTG 638
QY 234 TCAAGAGGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAAAGTGGTGA 293
DB 637 TCAAGAGGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAAAGTGGTGA 578
QY 294 AAATTACATCTGAATACCAAGACTCAGAGAAATGCAAGAGAGGCTGAACGATTCAATG 353
DB 577 AAATTACATCTGAATACCAAGACTCAGAGAAATGCAAGAGAGGCTGAACGATTCAATG 518
QY 354 TACTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAG 413
DB 517 TACTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAG 458
```

```
QY 414 TTCCAACAAAAGCTGTGTCATCTGATAACAAACCTATGTTTAACTTGGATAAGCTGAAGG 473
DB 457 TTCCAACAAAAGCTGTGTCATCTGATAACAAACCTATGTTTAACTTGGATAAGCTGAAGG 398
QY 474 AAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATG 533
DB 397 AAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATG 338
QY 534 AGAACTGAAAAGAGAGGAGCGGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGGAA 593
DB 337 AGAACTGAAAAGAGAGGAGCGGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGGAA 278
QY 594 CCACAGAGATACAGAGGCAAGAGAGCAAGAGAGCAAGAGCGCTTTGGGATTTGCCTGAT 653
DB 277 CCACAGAGATACAGAGGCAAGAGAGCAAGAGAGCAAGAGCGCTTTGGGATTTGCCTGAT 218
QY 654 GAAAAGTTCCTGATACCTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCT 713
DB 217 GAAAAGTTCCTGATACCTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCT 158
QY 714 ACATATATGCTTAATGACAGTCATGTGCTTACGTCCTGCTCGCAATGAGGAGCATG 773
DB 157 ACATATATGCTTAATGACAGTCATGTGCTTACGTCCTGCTCGCAATGAGGAGCATG 98
QY 774 TACCCAGAGTACATCCATCAACTGCGGAGCAGTGTGAGCTTATTGCTGTTTCAAGCTTTAA 833
DB 97 TACCCAGAGTACATCCATCAACTGCGGAGCAGTGTGAGCTTATTGCTGTTTCAAGCTTTAA 38
QY 834 GGTGTTGTGTTTTGTTTTGTTTGAATGATGTTGCTTGT 870
DB 37 GGTGTTGTGTTTTGTTTTGTTTGAATGATGTTGCTTGT 1
```

```
RESULT 13
BX456776
LOCUS BX456776 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004YA11 linear EST 06-MAY-2004
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION BX456776
VERSION BX456776.2 GI:47072683
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 843)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 22, 2003 this sequence version replaced gi:31036705.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 6407.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0CAP004AA06QPI&c=6407.r.
FEATURES
source
1..843
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP004YA11"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
ORIGIN
```


double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN	
Query Match	91.8%; Score 801.6; DB 5; Length 843;
Best Local Similarity	99.4%; Pred. No. 8.6e-200;
Mismatches	0; Mismatches 4; Indels 1; Gaps 1;
Matches	815; Conservative 0;
Qy 54 AGCTTGGCGAATAAAGCAAGATGTTCTGCTGCTGTTGGAGACCAAGGGAATAAGC 113	
Db 1 AGCTTGGCGAATAAAGCAAGATGTTCTGCTGCTGTTGGAGACCAAGGGAATAAGC 60	
Qy 114 AAGATCTTATCCACAGACTCCAGCATATCTTGAAGAACATGCTGAGAGGAGCAATG 173	
Db 61 AAGATCTTATCCACAGACTCCAGCATATCTTGAAGAACATGCTGAGAGGAGCAATG 120	
Qy 174 AAGAAGATGTACTGGGAGATGAAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTG 233	
Db 121 AAGAGATGTACTGGGAGATGAAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTG 180	
Qy 234 TCAAAGAGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGCTGA 293	
Db 181 TCAAAGAGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGCTGA 240	
Qy 294 AAATTACATCTGAAATACCAAGACTGAGAGATGCAAGAGAGGGCTGAACGATTCATG 353	
Db 241 AAATTACATCTGAAATACCAAGACTGAGAGATGCAAGAGAGGGCTGAACGATTCATG 300	
Qy 354 TACTGTGAGCTTGGAGTAAGAAAGCTGCTGGGAGCTAGTGTGGGATTTCTTCAG 413	
Db 301 TACTGTGAGCTTGGAGTAAGAAAGCTGCTGGGAGCTAGTGTGGGATTTCTTCAG 360	
Qy 414 TTCCAAACAAAGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAG 473	
Db 361 TTCCAAACAAAGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAG 420	
Qy 474 AAGAGCTCAAAGATTTGGTTTGAATGTCCTTCAATCTCCAGAAAGCTGGAAGATGATG 533	
Db 421 AAGAGCTCAAAGATTTGGTTTGAATGTCCTTCAATCTCCAGAAAGCTGGAAGATGATG 480	
Qy 534 AGAACTGAAAGAGAGAGGAGCGATTTGGATGTCACAGATTTGACAGTTGAGCTGGAA 593	
Db 481 AGAACTGAAAGAGAGAGGAGCGATTTGGATGTCACAGATTTGACAGTTGAGCTGGAA 540	
Qy 594 CCACAGAGGATACAGAGCGAAAGAGAGAAAGAGCAGAGCGCTTTGGGATTCCTGAT 653	
Db 541 CCACAGAGGATACAGAGCGAAAGAGAGAAAGAGCAGAGCGCTTTGGGATTCCTGAT 600	
Qy 654 GAAAAGTTCTGATACCTTCTGTTCTCCAGTGTTCATATTCCTCTCTTCTTCTGCTC 713	
Db 601 GAAAAGTTCTGATACCTTCTGTTCTCCAGTGTTCATATTCCTCTCTTCTTCTGCTC 660	
Qy 714 ACATATATGCTTAATGCAAGTCATGCTAGCTAGCTCCTGCTGCAATGAGGAGCATG 773	
Db 661 ACATATATGCTTAATGCAAGTCATGCTAGCTAGCTCCTGCTGCAATGAGGAGCATG 720	
Qy 774 TACCCAGGTACATCCATGAACTGCGCAGCAGGTTTGCATTTATGCTGTTTTCAGCTTAA 833	
Db 721 TACCCAGGTACATCCATGAACTGCGCAGCAGGTTTGCATTTATGCTGTTTTCAGCTTAA 779	
Qy 834 GGTGTTGTTGTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 873	
Db 780 GGTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 819	

RESULT 14
BU599301
LOCUS
DEFINITION
ACENCOURT_8908959 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6458323
5', mRNA sequence.
BU599301
ACCESSION
BU599301.1 GI:23251060

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

EST.
Homo sapiens (human)
Homo sapiens

FEATURES
source

1..823
/organism="Homo sapiens"
/mol type="mRNA"
/db xref="taxon:9606"
/clone="IMAGE:6458323"
/tissue_type="mixed (pool of 40 RNAs)"
/lab host="DH10B (T1-phage-resistant)"
/clone lib="NIH_MGC_142"
/notes="Vector: pDNR-LIB; Site 1: Sfil (ggccattatggcc); Site 2: Sfil (ggccctctggcc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGACGAGTGGCCATTACGCGCGG-3' and 5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH_MGC_142). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN	
Query Match	91.5%; Score 798.8; DB 5; Length 823;
Best Local Similarity	99.3%; Pred. No. 4.7e-199;
Mismatches	811; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
Qy 1 TGGAGTGGGGGTAAACAAGATGGCGACGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 60	
Db 8 TGGAGTGGGGGTAAACAAGATGGCGACGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 67	
Qy 61 CGAATAAAGCAAGATGTTCTGCTGCTGTTGGAGACCAAGGGAATAAGCAAGATCT 120	
Db 68 CGAATAAAGCAAGATGTTCTGCTGCTGTTGGAGACCAAGGGAATAAGCAAGATCT 127	
Qy 121 TATCCACAGACTCCAGCATATCTTGAAGAACATGCTGAGAGGAGGCAATGAGAGAGA 180	
Db 128 TATCCACAGACTCCAGCATATCTTGAAGAACATGCTGAGAGGAGGCAATGAGAGAGA 187	
Qy 181 TGTACTGGGAGATGAAACAGAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 240	
Db 188 TGTACTGGGAGATGAAACAGAGAGAGAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 247	
Qy 241 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTC 300	
Db 248 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTC 307	
Qy 301 ATCTGAAATACCAAGACTGAGAGATGCAAGAGAGGCTGAACGATTCAATGTACTCTGT 360	

Db	308	ATCTGAAATACACAGACTGAGAGAATGCAGAGAGAGGGCTGAACGATTCAAATGTACTGT	367
Qy	361	GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTTCAGTTCACAC	420
Db	368	GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTTCAGTTCACAC	427
Qy	421	AAAAGGTCCTCATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGAGC	480
Db	428	AAAAGGTCCTCATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGAGC	487
Qy	481	TCAAGAGTTTGGTTTGAATCTCTCTTCAATCTCCAGAAAGCTCTCAAGATCATGAGAACT	540
Db	488	TCAAGAGTTTGGTTTGAATCTCTCTTCAATCTCCAGAAAGCTCTCAAGATCATGAGAACT	547
Qy	541	GAAGAGAGGAAGGAGCGATTTGGGATTGTCAAAAGTTTCAGCTGGAAGCTGGAACCAAGA	600
Db	548	GAAGAGAGGAAGGAGCGATTTGGGATTGTCAAAAGTTTCAGCTGGAAGCTGGAACCAAGA	607
Qy	601	GGATACAGAGGCAAGAAAGAGGAGGAAAGAGCAGAGCGCTTTGGGATTCCTTGATGAAAAGT	660
Db	608	GGATACAGAGGCAAGAAAGAGGAGGAAAGAGCAGAGCGCTTTGGGATTCCTTGATGAAAAGT	667
Qy	661	TCCTGATACCTTCTGTTCTTCAGTGTCTTTCGATTTCTCTCTCTTCTTCTTGGTCAATATA	720
Db	668	TCCTGATACCTTCTGTTCTTCAGTGTCTTTCGATTTCTCTCTCTTCTTCTTGGTCAATATA	727
Qy	721	TGCTTAATGCACAGTCAATGCTGCTCAGTCTGCTGCTGCAATGAGGAGCATGTACCCCA	780
Db	728	TGCTTAATGCACAGTCAATGCTGCTCAGTCTGCTGCTGCAATGAGGAGCATGTACCCCA	787
Qy	781	GGTACATCCAT-GAACTGCGGACGAGCTTTGACTT	814
Db	788	GGTACATCCATGGAAGCTGCGGACGAGCTTTGACTT	822
RESULT 15			
BU508603			
LOCUS	BU508603	914 bp	linear
DEFINITION	AGENCOURT 10179614 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6502926		
ACCESSION	BU508603		
VERSION	BU508603.1		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	NH-MGC http://mhc.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L14M14059 Row: m Column: 07 High quality sequence stop: 662.		
FEATURES			
source	Location/Qualifiers		
	1. .914		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6502926"		
	/tissue_type="leiomyosarcoma"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_71"		
	/note="Organ: uterus; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr."		

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:18:12 ; Search time 560.261 Seconds
(without alignments)
9224.147 Million cell updates/sec

Title: US-09-788-476A-3
Perfect score: 873
Sequence: 1 tggagtggggtaacaaga.....tgattatgttctgttgaat 873

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001as.*
- 5: Geneseq2001bs.*
- 6: Geneseq2002as.*
- 7: Geneseq2002bs.*
- 8: Geneseq2003as.*
- 9: Geneseq2003bs.*
- 10: Geneseq2003cs.*
- 11: Geneseq2003ds.*
- 12: Geneseq2004as.*
- 13: Geneseq2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	873	100.0	873	6	ABX13936
2	873	100.0	894	6	ABX13935
3	873	100.0	1071	5	AAC88100
4	873	100.0	1154	5	AAS29109
5	873	100.0	1154	6	ABS68249
6	873	100.0	1154	10	ADL25243
7	873	100.0	1520	6	ABN59623
8	853.4	97.8	891	6	AAS62602
9	733.4	84.0	3147	10	ADI22609
10	690.8	79.1	2553	5	ADL62461
11	603	69.1	620	13	ACN39490
12	595.2	68.2	2553	5	ADL62461
13	498.4	57.1	1022	10	ADI22423
14	465.6	53.3	471	3	AAC00738
15	451.6	51.7	558	12	ADP28791
16	336.2	38.5	463	9	ACH28217
17	334	38.3	408	5	AH81555
18	265.4	30.4	5469	4	AAK84119
19	265.4	30.4	5469	4	AAI37635
20	265.4	30.4	5469	4	AAI06683

C	21	265.4	30.4	5469	4	ABA08022	Human ova
C	22	265.4	30.4	5469	8	ABX60623	CDNA enco
C	23	265.4	30.4	5469	12	ADJ31373	Human mus
C	24	265.4	30.4	9453	4	AAK84118	Human imm
C	25	265.4	30.4	9453	4	AAI37634	Human mus
C	26	265.4	30.4	9453	4	AAI06682	Human rep
C	27	265.4	30.4	9453	4	ABA08021	Human ova
C	28	265.4	30.4	9453	8	ABX60622	CDNA enco
C	29	265.4	30.4	9453	12	ADJ31372	Human mus
C	30	258.6	29.6	266	2	AAV89379	EST clone
C	31	246.2	28.2	300	10	ADI22386	Rat liver
C	32	227.2	26.0	591	2	AAK85623	Novel cDN
C	33	211.8	24.3	255	5	AAH82022	Rat diffe
C	34	136.8	22.5	232	8	AAH9101	Mouse DST
C	35	164.8	18.9	399	10	ADH50756	Primary r
C	36	154.2	17.7	215	6	ABL41869	Novel mur
C	37	92.2	10.6	463	5	ADL41869	Human ova
C	38	56	6.4	208765	12	ADQ97430	Mouse can
C	39	55.4	6.3	1686	2	AAQ87587	DNA enco
C	40	55	6.3	399	5	AAS90682	DNA enco
C	41	55	6.3	399	5	AAS93406	DNA enco
C	42	54.4	6.2	1997	13	ADR07778	Full leng
C	43	54	6.2	5659	6	ABL32375	Human imm
C	44	54	6.2	5659	6	ABL34487	Human met
C	45	54	6.2	5659	7	ADS99748	Complemen

ALIGNMENTS

RESULT 1
ABX13936
ID ABX13936 standard; DNA; 873 BP.

XX AC ABX13936;

DT 03-MAR-2003 (first entry)

XX DNA encoding novel human protein HCC-1, PCR extended form.

Human; hcc-1; human hepatocellular carcinoma tissue; HCC-M cell line;
cancer; hepatocellular carcinoma; antisense gene therapy; gene; ds.

OS Homo sapiens.

XX Key Location/Qualifiers
FH CDS 20...652
FT /*tag= a
FT /product= "HCC-1"

XX US2002107190-A1.

XX 08-AUG-2002.

XX 21-FEB-2001; 2001US-00788476.

XX 25-FEB-2000; 2000US-0185116P.

XX (CHUN/) CHUNG C M.

XX (CHAN/) CHAN L.

XX (OUKK/) OU K.

XX (ONGS/) ONG S.

XX (SEOW/) SEOW T K.

XX (LIAN/) LIANG C R.

XX (CHOO/) CHOONG M L.

XX (TANL/) TAN L K.

PI Chung CM, Chan L, Ou K, Ong S, Seow TK, Liang CR, Choong ML;

XX Tan LK;

DR WPI; 2002-697878/75.

XX New nucleic acid which is differentially expressed in human

60/185,116

		Matches	873;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	1	1	1	1	1	1	1	1	1	1
Db	6	6	6	6	6	6	6	6	6	6	6
Qy	61	61	61	61	61	61	61	61	61	61	61
Db	66	66	66	66	66	66	66	66	66	66	66
Qy	121	121	121	121	121	121	121	121	121	121	121
Db	126	126	126	126	126	126	126	126	126	126	126
Qy	181	181	181	181	181	181	181	181	181	181	181
Db	186	186	186	186	186	186	186	186	186	186	186
Qy	241	241	241	241	241	241	241	241	241	241	241
Db	246	246	246	246	246	246	246	246	246	246	246
Qy	301	301	301	301	301	301	301	301	301	301	301
Db	306	306	306	306	306	306	306	306	306	306	306
Qy	361	361	361	361	361	361	361	361	361	361	361
Db	366	366	366	366	366	366	366	366	366	366	366
Qy	421	421	421	421	421	421	421	421	421	421	421
Db	426	426	426	426	426	426	426	426	426	426	426
Qy	481	481	481	481	481	481	481	481	481	481	481
Db	486	486	486	486	486	486	486	486	486	486	486
Qy	541	541	541	541	541	541	541	541	541	541	541
Db	546	546	546	546	546	546	546	546	546	546	546
Qy	601	601	601	601	601	601	601	601	601	601	601
Db	606	606	606	606	606	606	606	606	606	606	606
Qy	661	661	661	661	661	661	661	661	661	661	661
Db	666	666	666	666	666	666	666	666	666	666	666
Qy	721	721	721	721	721	721	721	721	721	721	721
Db	726	726	726	726	726	726	726	726	726	726	726
Qy	781	781	781	781	781	781	781	781	781	781	781
Db	786	786	786	786	786	786	786	786	786	786	786
Qy	841	841	841	841	841	841	841	841	841	841	841
Db	846	846	846	846	846	846	846	846	846	846	846

RESULT 3

AAC88100

ID AAC88100 standard; cDNA; 1071 BP.

XX

AC AAC88100;

XX

DT 09-MAR-2001 (first entry)

XX

DE Human FLEXHT-31 nucleotide sequence SEQ ID NO:86.

XX

KW	Human; FLEXHT; full-length molecules expressed in human tissue;
KW	diagnosis; gene expression; genetic linkage; genetic variability;
KW	antianemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;
KW	cytostatic; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;
KW	anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;
KW	antipsoriatic; antirheumatic; antiulcer; gene therapy; anaemia; gout;
KW	epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;
KW	cancer; immunological disorder; asthma; bronchitis; cirrhosis;
KW	Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;
KW	osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;
KW	ulcerative colitis; ps.
OS	Homo sapiens.
XX	WO200070047-A2.
XX	23-NOV-2000.
XX	12-MAY-2000; 2000WO-US013299.
XX	14-MAY-1999; 99US-00311894.
PR	14-MAY-1999; 99US-00311937.
PR	14-MAY-1999; 99US-00311940.
XX	(INCY-) INCYTE GENOMICS INC.
XX	Yue H, Tang YT, Lal P, Reddy R, Batra S, Baughn MR, Yang J;
PI	Azimzai Y, Lu DAM, Au-Young J, Shih LL;
XX	WPI: 2001-016234/02.
DR	P-PSDB; AAB36609.
XX	Human FLEXHT protein and DNA sequences, useful for treating immunological
PT	disorders, developmental disorders, and cancers.
XX	Claim 5; Page 154; 168pp; English.
XX	AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules
CC	expressed in human tissues) proteins given in AAB36579 to AAB36633. The
CC	present invention describes an isolated polypeptide (A) comprising an
CC	amino acid sequence selected from one of 55 amino acid sequences 42-876
CC	residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %
CC	identical sequence, and a biologically active or immunogenic fragment of
CC	the sequence. The FLEXHT proteins can have antianemic, anticonvulsant,
CC	antiarteriosclerotic, immunomodulatory, cytotostatic, antiasthmatic,
CC	antiinflammatory, hepatotropic, antidiabetic, anti-gout, antithyroid,
CC	neuroprotective, antiarthritic, osteopathic, antipsoriatic, antiulcer and
CC	antirheumatic activities, and can be used in gene therapy. The
CC	polynucleotide sequences can be used to express the protein sequences.
CC	Pharmaceutical compositions comprising FLEXHT can be used to treat
CC	diseases or conditions associated with altered expression of functional
CC	FLEXHT. The proteins and polynucleotides can be used to diagnose and
CC	treat disorders including anaemia, epilepsy, arteriosclerosis,
CC	atherosclerosis, developmental disorders, cancers, and immunological
CC	disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,
CC	diabetes mellitus, gout, Grave's disease, multiple sclerosis,
CC	osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and
CC	ulcerative colitis
XX	Sequence 1071 BP; 338 A; 190 C; 297 G; 246 T; 0 U; 0 Other;

Query Match 100.0%; Score 873; DB 5; Length 1071;

Best Local Similarity 100.0%; Pred. No. 4.4e-232;

Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGAGTGGGGTAAACAAGATGGCGACGAGCGTGGAGCTCCATAGCTAAAGCTTGC 60

Db 165 TGGAGTGGGGTAAACAAGATGGCGACGAGCGTGGAGCTCCATAGCTAAAGCTTGC 224

Qy 61 CGAACTAAGCAAGATGCTTCTGCTGTTTGGAGACCAAGGGAATAAGCAAGATCT 120

Db 225 CGAACTAAGCAAGATGCTTCTGCTGTTTGGAGACCAAGGGAATAAGCAAGATCT 284

QY 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGCGCAAAATGAAGA 180
Db 285 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGCGCAAAATGAAGA 344
QY 181 TGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 240
Db 345 TGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 404
QY 241 GGAAGAACCCCTGAAACAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAANAATTAC 300
Db 405 GGAAGAACCCCTGAAACAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAANAATTAC 464
QY 301 ATCTGAAATACACAGACTGAGAGAAATGCAAGAGAGGCTGGAACGATTCATGTTACTCTGT 360
Db 465 ATCTGAAATACACAGACTGAGAGAAATGCAAGAGAGGCTGGAACGATTCATGTTACTCTGT 524
QY 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAAC 420
Db 525 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAAC 584
QY 421 AAAAGTCTGTCTATCTGATTAACAACTATGGTTAACTTGGATAAGCTGAAGAAAGAGC 480
Db 585 AAAAGTCTGTCTATCTGATTAACAACTATGGTTAACTTGGATAAGCTGAAGAAAGAGC 644
QY 481 TCAAAGATTTGGTTGAACTGCTCTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACT 540
Db 645 TCAAAGATTTGGTTGAACTGCTCTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACT 704
QY 541 GAAAGAGGAGGAGGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACTGGAAACCACAGA 600
Db 705 GAAAGAGGAGGAGGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACTGGAAACCACAGA 764
QY 601 GGATACAGAGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
Db 765 GGATACAGAGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 824
QY 661 TCCTGATCTTCTGTTCTCCAGTGTTCCTCAATTTCTCTCTTCTTCTTCTTCTTCTTCTT 720
Db 825 TCCTGATCTTCTGTTCTCCAGTGTTCCTCAATTTCTCTCTTCTTCTTCTTCTTCTTCTT 884
QY 721 TGCTTAAATGCACAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 885 TGCTTAAATGCACAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
QY 781 GGTACATCCATGAAGTGGCAGAGGTTGACATTAATGCTGTTTACGTTTAAAGTTGTT 840
Db 945 GGTACATCCATGAAGTGGCAGAGGTTGACATTAATGCTGTTTACGTTTAAAGTTGTT 1004
QY 841 GTGTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 873
Db 1005 GTGTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1037

RESULT 4
ID AAS29109 standard; cdna; 1154 BP.
XX AC AAS29109;
XX AC AAS29109;
DT 21-NOV-2001 (first entry)
DE cDNA encoding for human DNA-binding protein #80.
XX Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein; DNA organisation;
KW gene transcription; malignant disease; autoimmune disorder;
KW rheumatic disease; genetic abnormality; infectious disease;
KW neurological disorder; gene therapy; immunomodulatory; anti-HIV;
XX anti rheumatic; anti microbial; cytostatic; ss.
OS Homo sapiens.
XX WO200155162-A1.
PN

XX 02-AUG-2001.
PD 17-JAN-2001; 2001WO-US001305.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184654P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.

not entered
60 lines, 005
Not entered
checked here

PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254037P.
PR 05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX
XX

DR WPI; 2001-465557/50.
XX P-PSDB; AAU18233.
PT Nucleic acid molecules encoding human secreted chromosomal binding
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers.
XX
XX Claim 4; SEQ ID NO 90; 561pp; English.
XX
CC The present invention relates to the isolation of novel DNA-binding
CC proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding for
CC these proteins. DNA-binding proteins such as histones, chromo (Chromatin
CC organisation modifier) domain proteins, and Y-box binding proteins may
CC contribute to diseases resulting from aberrant DNA organisation and/or
CC gene transcription. The sequences of the invention are useful in
CC screening assays to identify antagonists and/or agonists that may enhance
CC or block activities mediated by DNA-binding proteins. Blockers of DNA-
CC binding proteins may be useful in treating disorders such as malignant
CC diseases (e.g. cancer), autoimmune disorders (e.g. diabetes mellitus),
CC rheumatic diseases (e.g. rheumatoid arthritis), genetic abnormalities
CC (e.g. cystic fibrosis), infectious diseases (e.g. HIV) and neurological
CC disorders (e.g. Alzheimer's disease). The polynucleotide sequences of the
CC invention may also be used in gene therapy. AAS29030-AAS29157 represent
CC cDNA sequences encoding for novel DNA-binding proteins. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1154 BP; 394 A; 186 C; 291 G; 281 T; 0 U; 2 Other;
Query Match 100.0%; Score 873; DB 5; Length 1154;
Best Local Similarity 100.0%; Pred. No. 4.5e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGAGTGGGGTAAACAAGATGCGACCGAGCGTGGAGCTCCATAAGCTAAAGCTTGC 60
DB 182 TGGAGTGGGGGTAAACAAGATGCGACCGAGCGTGGAGCTCCATAAGCTAAAGCTTGC 241
QY 61 CGAACTAAAGCAAGAATGCTTCTCGTGGTGGAGCAAGGGAATAAGCAAGATCT 120
DB 242 CGAACTAAAGCAAGAATGCTTCTCGTGGTGGAGCAAGGGAATAAGCAAGATCT 301
QY 121 TATCCACAGACTCAGGCATATCTTGAAGAACAATGCTGAAGAGGAGGCAATGAAGAAGA 180
DB 302 TATCCACAGACTCAGGCATATCTTGAAGAACAATGCTGAAGAGGAGGCAATGAAGAAGA 361
QY 181 TGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 240
DB 362 TGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 421
QY 241 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAGAAAGTGTGAAATTTAC 300
DB 422 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAGAAAGTGTGAAATTTAC 481
QY 301 ATCTGAAATACACAGACTCAGAGATGCGAGAGGGCTGAAGGATTAATGATCTGT 360
DB 482 ATCTGAAATACACAGACTCAGAGATGCGAGAGGGCTGAAGGATTAATGATCTGT 541
QY 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTGGGATTTCTTCAGTTCCAAC 420
DB 542 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTGGGATTTCTTCAGTTCCAAC 601
QY 421 AAAAGGTCTGTCATCTGATAACAAACCTATGTTTAACTTGGATAAGCTGAAGGAAAGAGC 480
DB 602 AAAAGGTCTGTCATCTGATAACAAACCTATGTTTAACTTGGATAAGCTGAAGGAAAGAGC 661
QY 481 TCAAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACT 540
DB 662 TCAAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACT 721
QY 541 GAAAAGAGGAGGAGCGATTTGGGATTTGTCACAAAGTTGACGTTGGAATGGAACCAAGAGC 600
DB 722 GAAAAGAGGAGGAGCGATTTGGGATTTGTCACAAAGTTGACGTTGGAATGGAACCAAGAGC 781


```
Qy 1 TGAAGTGAAGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 60
Db 182 TGAAGTGAAGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 241
Qy 61 CGAATTAAGCAGAGATGCTTCTCGTGGTGGAGACCAAGGGAATAAGCAAGATCT 120
Db 242 CGAATTAAGCAGAGATGCTTCTCGTGGTGGAGACCAAGGGAATAAGCAAGATCT 301
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATCTGTGAAGAGGAGGCAATGAAGAAGA 180
Db 302 TATCCACAGACTCCAGGCATATCTTGAAGAACATCTGTGAAGAGGAGGCAATGAAGAAGA 361
Qy 181 TGTACTGGAGATGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 362 TGTACTGGAGATGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
Qy 241 GGAAGACCCCTGAAAGAACTCTGATGTGGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 422 GGAAGAAACCCCTGAAAGAACTCTGATGTGGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 481
Qy 301 ATCTGAAATACCAAGACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 482 ATCTGAAATACCAAGACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541
Qy 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGAGAGCTAGGTTGGAGTTCTTCAGTTCAAC 420
Db 542 GAGCTTGGAGAGTAAGAAAGCTGCTCGGAGAGCTAGGTTGGAGTTCTTCAGTTCAAC 601
Qy 421 AAAAGGTCTGATCTGATTAACAAACCTATGGTTAACTTGGTAAGCTGAAGAGAGAGAG 480
Db 602 AAAAGGTCTGATCTGATTAACAAACCTATGGTTAACTTGGTAAGCTGAAGAGAGAGAG 661
Qy 481 TCAAGATTTGGTTGATCTCTTCAATCTCAGAAAGCTCTGAAGATGATGAGAACT 540
Db 662 TCAAGATTTGGTTGATCTCTTCAATCTCAGAAAGCTCTGAAGATGATGAGAACT 721
Qy 541 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Db 722 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781
Qy 601 GGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 782 GGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
Qy 661 TCCTGATACCTTCTGTTCTCAGTGTGTTTCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 842 TCCTGATACCTTCTGTTCTCAGTGTGTTTCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 901
Qy 721 TGCTTAATGCACAGTCATGTGCTACGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 902 TGCTTAATGCACAGTCATGTGCTACGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 961
Qy 781 GGTACATCCATGAATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 962 GGTACATCCATGAATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
Qy 841 GTGTTTTGTTTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 873
Db 1022 GTGTTTTGTTTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1054
```

RESULT 6

ADC25243

ID ADC25243 standard; cDNA; 1154 BP.

XX

XX

AC ADC25243;

XX

DT 18-DEC-2003 (first entry)

XX

DE Human cDNA from extracellular matrix gene 80.

XX

KW Extracellular matrix protein; cytostatic; antibacterial; virucide;

KW neuroprotective; gynaecological; gastrointestinal; cardiant;

KW

cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen; respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic; KW nontropic; antiallergic; cancer; bacterial infection; viral infection; KW neural disorder; immune system disorder; blood disorder; KW muscular disorder; reproductive disorder; gastrointestinal disorder; KW pulmonary disorder; cardiovascular disorder; renal disorder; KW inflammatory disorder; proliferative disorder; human; gene therapy; ss; gene.

Homo sapiens.

US2003049650-A1.

13-MAR-2003.

07-MAR-2002; 2002US-00091483.

31-JAN-2000; 2000US-0179065P.

04-FEB-2000; 2000US-0180828P.

24-FEB-2000; 2000US-0184664P.

02-MAR-2000; 2000US-0186350P.

16-MAR-2000; 2000US-0189874P.

17-MAR-2000; 2000US-0190076P.

18-APR-2000; 2000US-0198123P.

19-MAY-2000; 2000US-0205515P.

07-JUN-2000; 2000US-0209467P.

28-JUN-2000; 2000US-0214886P.

30-JUN-2000; 2000US-0215135P.

07-JUL-2000; 2000US-0216647P.

07-JUL-2000; 2000US-0216880P.

11-JUL-2000; 2000US-0217487P.

14-JUL-2000; 2000US-0218290P.

26-JUL-2000; 2000US-0220963P.

26-JUL-2000; 2000US-0220964P.

14-AUG-2000; 2000US-0224518P.

14-AUG-2000; 2000US-0224519P.

14-AUG-2000; 2000US-0225213P.

14-AUG-2000; 2000US-0225214P.

14-AUG-2000; 2000US-0225366P.

14-AUG-2000; 2000US-0225467P.

14-AUG-2000; 2000US-0225468P.

14-AUG-2000; 2000US-0225270P.

14-AUG-2000; 2000US-0225447P.

14-AUG-2000; 2000US-0225757P.

14-AUG-2000; 2000US-0225758P.

18-AUG-2000; 2000US-0225759P.

22-AUG-2000; 2000US-0226681P.

22-AUG-2000; 2000US-0226688P.

22-AUG-2000; 2000US-0227182P.

23-AUG-2000; 2000US-0227009P.

30-AUG-2000; 2000US-0228924P.

01-SEP-2000; 2000US-0229287P.

01-SEP-2000; 2000US-0229343P.

01-SEP-2000; 2000US-0229344P.

01-SEP-2000; 2000US-0229345P.

05-SEP-2000; 2000US-0229509P.

05-SEP-2000; 2000US-0229513P.

06-SEP-2000; 2000US-0230437P.

06-SEP-2000; 2000US-0230438P.

08-SEP-2000; 2000US-0231242P.

08-SEP-2000; 2000US-0231243P.

08-SEP-2000; 2000US-0231244P.

08-SEP-2000; 2000US-0231413P.

08-SEP-2000; 2000US-0231414P.

08-SEP-2000; 2000US-0232080P.

08-SEP-2000; 2000US-0232081P.

12-SEP-2000; 2000US-0231968P.

14-SEP-2000; 2000US-0232397P.

14-SEP-2000; 2000US-0232398P.

14-SEP-2000; 2000US-0232399P.

14-SEP-2000; 2000US-0232400P.

PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.

PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-0076484E.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Barash SC;
XX WPI; 2003-605749/57.
XX F-PSDB; ADC25371.
XX New DNA-binding proteins and gene encoding them, useful for diagnosing,
PT treating and/or preventing e.g. neurological, inflammatory, infectious,
PT cardiovascular, autoimmune, respiratory, neoplastic or digestive
PT diseases.
XX
PS Claim 1; SEQ ID NO 90; 226pp; English.
XX The invention relates to an isolated nucleic acid molecule (cDNA)
CC encoding a human extracellular matrix protein, representing one of 161
CC novel genes. Also included are recombinant vectors, host cells
CC (expressing the protein), the extracellular matrix proteins (including
CC their fragments, epitopes and homologues), an isolated antibody that
CC binds specifically to the protein, diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or absence of a mutation in the nucleic acid and diagnosing a
CC condition based on the presence or absence of the mutation), diagnosing a
CC pathological condition or susceptibility to a pathological condition
CC (comprising determining the presence or amount of expression of the
CC protein in a biological sample and diagnosing a condition based on the
CC presence or amount of expression of the protein), preventing, treating or
CC ameliorating a medical condition by administering the nucleic acid or
CC protein to a mammalian subject, identifying a binding partner to the
CC protein, the gene corresponding to the cDNA sequence, and identifying an
CC activity in a biological assay (comprising expressing the nucleic acid in
CC a cell, isolating the supernatant, detecting an activity in a biological
CC assay and identifying the protein in the supernatant having the
CC activity). The nucleic acids and proteins display the following
CC activities Cytostatic, antibacterial, Virucide, Neuroprotective,

Query Match 100.0%; Score 873; DB 10; Length 1154;
Best Local Similarity 100.0%; Pred. No. 4.5e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGTGGGGGTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 60
DB 182 TGGAGTGGGGGTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTGC 241
QY 61 CGAACTAAAGCAAGATGCTTCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATCT 120
DB 242 CGAACTAAAGCAAGATGCTTCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATCT 301
QY 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCGCAATGAAGA 180
DB 302 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCGCAATGAAGA 361
QY 181 TGTACTGGGAGATGAACAG 240
DB 362 TGTACTGGGAGATGAACAG 421
QY 241 GGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAGAGAGAGAGAGAGAG 300
DB 422 GGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAGAGAGAGAGAGAGAG 481
QY 301 ATCTGAAATACCAAGACTGAGAGAGATGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 482 ATCTGAAATACCAAGACTGAGAGAGATGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541

Qy 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 420
Db 542 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 601
Qy 421 AAAAGGCTCTCATCTGATACAAACCTATGTTTAACTTCGATAAGCTGAAGGAAGAGC 480
Db 602 AAAAGGCTCTCATCTGATACAAACCTATGTTTAACTTCGATAAGCTGAAGGAAGAGC 661
Qy 481 TCAAAGATTTGGTTGGAATGCTCTTCAATCTCCAGAAAGCTGGAAGATGATGAGAACT 540
Db 662 TCAAAGATTTGGTTGGAATGCTCTTCAATCTCCAGAAAGCTGGAAGATGATGAGAACT 721
Qy 541 GAAAAGAGGAAGAGCGATTTGGATTGTGCAAGTTTCAGCTGGAACCTGGAACCAAGA 600
Db 722 GAAAAGAGGAAGAGCGATTTGGATTGTGCAAGTTTCAGCTGGAACCTGGAACCAAGA 781
Qy 601 GGATACAGAGCAAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 782 GGATACAGAGCAAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
Qy 661 TCCTGATACATTTCTGTTCTCCAGTGTGTTTTCATTTCTCTCTCTCTCTCTCTCTCT 720
Db 842 TCCTGATACATTTCTGTTCTCCAGTGTGTTTTCATTTCTCTCTCTCTCTCTCTCTCT 901
Qy 721 TGCTTAATGCACAGTCATGTGCTACGTCTGCTCGCTCGCAATGAGGAGAGATGTACCCCA 780
Db 902 TGCTTAATGCACAGTCATGTGCTACGTCTGCTCGCTCGCAATGAGGAGAGATGTACCCCA 961
Qy 781 GGTACATCCATGAACCTCGGCGAGAGTTGACTTATTTGCTGTTTACGCTTTAAGTTGTT 840
Db 962 GGTACATCCATGAACCTCGGCGAGAGTTGACTTATTTGCTGTTTACGCTTTAAGTTGTT 1021
Qy 841 GTGTTTTGTTTTGATTATGTTGCTGTTAAT 873
Db 1022 GTGTTTTGTTTTGATTATGTTGCTGTTAAT 1054

RESULT 7

ABN59623
ID ABN59623 standard; cDNA; 1520 BP.
AC AC
XX ABN59623;
DT 28-JUN-2002 (first entry)
XX Novel human coding sequence SEQ ID NO: 34.
DE Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag; gene; ss.
XX Homo sapiens.
OS OS
XX WO200222660-A2.
PN 21-MAR-2002.
PD 10-SEP-2001; 2001WO-US026015.
XX 11-SEP-2000; 2000US-00659671.
PR (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
DR P-PSDB; ABB97210.
XX An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX

PS Claim 1; SEQ ID NO 34; 509pp; English.
XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a coding sequence of the
CC invention
XX
Qy Sequence 1520 BP; 442 A; 310 C; 423 G; 345 T; 0 U; 0 Other;
XX
Query Match 100.0%; Score 873; DB 6; Length 1520;
Best Local Similarity 100.0%; Pred. No. 5.1e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGGAGTGCAGGGGTAAACAAGATGCGACCGAGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 60
Db 617 TGGAGTGCAGGGGTAAACAAGATGCGACCGAGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 676
Qy 61 CGAACTAAAGCAAGAATGCTTCTGCTGCTGTTTGGAGCAAGGGAATAAGCAAGATCT 120
Db 677 CGAACTAAAGCAAGAATGCTTCTGCTGCTGTTTGGAGCAAGGGAATAAGCAAGATCT 736
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACAATGCTGAAGAGGAGCAATGAAGAAGA 180
Db 737 TATCCACAGACTCCAGGCATATCTTGAAGAACAATGCTGAAGAGGAGCAATGAAGAAGA 796
Qy 181 TGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAGCCCATTTGAGCTCCTCTGTCGAAGA 240
Db 797 TGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAGCCCATTTGAGCTCCTCTGTCGAAGA 856
Qy 241 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGAGTGGTGAATAATTAC 300
Db 857 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGAGTGGTGAATAATTAC 916
Qy 301 ATCTGAAATACACAGACTCAGAGAAATGCAAGAGAGGGCTGAACGATTCATGTACCTGT 360
Db 917 ATCTGAAATACACAGACTCAGAGAAATGCAAGAGAGGGCTGAACGATTCATGTACCTGT 976
Qy 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 420
Db 977 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 1036
Qy 421 AAAAGGCTCTCATCTGATACAAACCTATGTTTAACTTGGATAAGCTGAAGGAAGAGAGC 480
Db 1037 AAAAGGCTCTCATCTGATACAAACCTATGTTTAACTTGGATAAGCTGAAGGAAGAGAGC 1096
Qy 481 TCAAAGATTTGGTTGGAATGCTCTTCAATCTCCAGAAAGCTCTGAAGATGATGAGAACT 540
Db 1097 TCAAAGATTTGGTTGGAATGCTCTTCAATCTCCAGAAAGCTCTGAAGATGATGAGAACT 1156
Qy 541 GAAAAGAGGAAGAGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACCAAGA 600
Db 1157 GAAAAGAGGAAGAGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACCAAGA 1216
Qy 601 GGATACAGAGCAAGAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 1217 GGATACAGAGCAAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1276
Qy 661 TCCTGATACATTTCTGTTCTCCAGTGTGTTTTCATTTCTCTCTCTCTCTCTCTCTCT 720
Db 1277 TCCTGATACATTTCTGTTCTCCAGTGTGTTTTCATTTCTCTCTCTCTCTCTCTCTCT 1336
Qy 721 TGCTTAATGCACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGAGATGTACCCCA 780
Db 1337 TGCTTAATGCACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGAGATGTACCCCA 1396
Qy 781 GGTACATCCATGAACCTCGGCGAGAGTTTGAATTTGCTGTTTACGCTTTAAGTTGTT 840

Db	1397	GGTACATCCATGAAC	CTGGGACAGTTGTACTTATTGCTGCTGTTT	CAGCTTTAAGGTTGTT	1456
Qy	841	GTGTTTTTGTGTTT	GATTATATGCTTGTGTTAAT	873	
Db	1457	GTGTTTTTGTGTTT	GATTATGTTGCTTGTGTTAAT	1489	
RESULT 8					
AAS62602/c					
ID	AAS62602	standard;	cdna;	891 BP.	
AC	AAS62602;				
XX					
XX	14-FEB-2002	(first entry)			
XX					
DE	cdna	sequence #389	encoding novel human secreted protein.		
XX					
KW	Human secreted protein;	hyperproliferative disorder;	autoimmune disorder		
KW	immune deficiency disorder;	blood disorder;	inflammatory disorder;		
KW	infectious disorder;	gene therapy;	antimicrobial; hepatotropic;		
KW	immunosuppressive;	antirheumatic;	ss.		
XX					
OS	Homo sapiens.				
XX					
PN	WO200177291-A2.				
XX					
PD	18-OCT-2001.				
XX					
PF	29-MAR-2001;	2001WO-US010485.			
XX					
PR	06-APR-2000;	2000US-0195604P.			
XX					
XX	(GEMY)	GENETICS INST INC.			
PA					
XX					
PI	Wong GG, Clark HF, Pechtel K, Agostino MJ, Howes SH, Resnick RJ;				
PI	Gulukota K, Graham JR;				
XX					
DR	WPI; 2002-010900/01.				
XX					
PT	New polynucleotides encoding secreted proteins useful for treating e.g.				
PT	asthma, HIV and Crohn's disease.				
XX					
PS	Claim 1; Page 280; 391pp; English.				
XX					
CC	The present invention relates to the isolation of novel cdna sequences				
CC	which encode human secreted proteins. The cdna sequences have been				
CC	derived from a variety of human tissues. The invention also provides a				
CC	method for producing proteins from these polynucleotide sequences. The				
CC	proteins are useful for identifying compounds that modulate their				
CC	activity and production, and the cell is also useful for identifying				
CC	compounds that modulate expression of the polynucleotide sequences				
CC	encoding the secreted proteins. The sequences of the invention are useful				
CC	for treating diseases such as hyperproliferative disorders (e.g. cancer),				
CC	immune deficiency disorders (e.g. severe combined immunodeficiency				
CC	(SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders				
CC	(e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and				
CC	infectious disorders (e.g. hepatitis). The polynucleotide sequences of				
CC	the invention are also useful in gene therapy. AAS62214-AAS62838				
CC	represent the cdna sequences of the invention that encode for novel human				
CC	secreted proteins				
XX					
SQ	Sequence 891 BP; 221 A; 224 C; 157 G; 289 T; 0 U; 0 Other;				
Query Match		97.8%;	Score 853.4;	DB 6; Length 891;	
Best Local Similarity		99.8%;	Pred. No. 1.1e-226;		
Matches 865; Conservative		0; Mismatches	1; Indels	1; Gaps	1;
Qy	8	AGGGTTAACAGATGGCGACCGACGCGTGGAGCTCCATAGCTTAAGCTTGCCGA	ACTA	67	
Db	886	AAGGGTTAACAGATGGCGACCGACGCGTGGAGCTCCATAGCTTAAGCTTGCCGA	ACTA	827	
Qy	68	AAGCAAGAAATGCTTGTGCTGCTGGTTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCAC	127		

CC marker corresponds to a secreted protein or to a transcribed
CC polynucleotide or its portion. The level of expression of the marker is
CC assessed by detecting the presence in the sample, a protein or protein
CC fragment corresponding to the marker. The presence of protein or protein
CC fragment is detected using an antibody that specifically binds with the
CC protein or protein fragment. Alternatively, the level of expression of
CC the marker is assessed by detecting the presence of a transcribed
CC polynucleotide which anneals with the marker or anneals with a portion of
CC the polynucleotide comprising the marker, under stringent conditions. The
CC marker is also used for monitoring the progression of ovarian cancer in a
CC patient which involves detecting expression of the marker in a patient
CC sample at a first point in time, repeating the method at a subsequent
CC time and comparing the level of expression. The method is carried out
CC using an ovarian tissue sample. A composition comprising a marker,
CC polypeptide or antibody of the invention is used to treat ovarian cancer.
CC This sequence represents a human ovarian cancer DNA marker of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2553 BP; 690 A; 597 C; 570 G; 696 T; 0 U; 0 Other;

Query Match 79.1%; Score 690.8; DB 5; Length 2553;
Best Local Similarity 99.7%; Pred. No. 3e-181;
Matches 692; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 93 TGGAGACCAAGGATTAAGCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAC 152
DB |||||
QY 715 TGGAGACCAAGGATTAAGCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAC 774
DB |||||
QY 153 ATGCTGAAGAGGAGGCAAAATGAAGAGATGTACTGGGAGATGAAACAGAGGAAGAGAAA 212
DB |||||
QY 775 ATGCTGAAGAGGAGGCAAAATGAAGAGATGTACTGGGAGATGAAACAGAGGAAGAGAAA 834
DB |||||
QY 213 CAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAGAACTGTTGATGG 272
DB |||||
QY 835 CAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGAACCCCTGAAAGAACTGTTGATGG 894
DB |||||
QY 273 CAGCAGAGAGAAAGTGGTGAATAATACATCTGAATACACAGACTGAGAGAAATGCGAGA 332
DB |||||
QY 895 CAGCAGAGAGAAAGTGGTGAATAATACATCTGAATACACAGACTGAGAGAAATGCGAGA 954
DB |||||
QY 333 AGAGGGCTGAACGATTCATGTTGAGTCTGAGAGATGAAGAGCTGCTCGGSCAG 392
DB |||||
QY 955 AGAGGGCTGAACGATTCATGTTGAGTCTGAGAGATGAAGAGCTGCTCGGSCAG 1014
DB |||||
QY 393 CTAGGTTTGGGATTTCTTCAAGTTCCAAACAAAGCTCTGTCATCTGATAACAAACCTATGG 452
DB |||||
QY 1015 CTAGGTTTGGGATTTCTTCAAGTTCCAAACAAAGCTCTGTCATCTGATAACAAACCTATGG 1074
DB |||||
QY 453 TTAACCTTGGATTAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAATGCTCTTCAATCT 512
DB |||||
QY 1075 TTAACCTTGGATTAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAATGCTCTTCAATCT 1134
DB |||||
QY 513 CCAGAAAGCTGGAAGATGATGAGAACTGAAAGAGGAAGGAGCGATTTGGGATTTGCA 572
DB |||||
QY 1135 CCAGAAAGCTGGAAGATGATGAGAACTGAAAGAGGAAGGAGCGATTTGGGATTTGCA 1194
DB |||||
QY 573 CAAGTTCCAGCTGGAACCTGGAACCAACAGAGGATACAGAGGCAAGAGGAAAGAGAGCAG 632
DB |||||
QY 1195 CAAGTTCCAGCTGGAACCTGGAACCAACAGAGGATACAGAGGCAAGAGGAAAGAGCAG 1254
DB |||||
QY 633 AGGCTTTGGGATTTGCTGATGAAGAAAGTTCTGATACCTTTCTGTTCTCCAGTGTTCCTCA 692
DB |||||
QY 1255 AGGCTTTGGGATTTGCTGATGAAGAGTTCTGATACCTTTCTGTTCTCCAGTGTTCCTCA 1314
DB |||||
QY 693 TTTCTCTCTCTTCTTCTTGGTTCACATATATGCTTAATGACAGTTCATGTCCTACGTCCT 752
DB |||||
QY 1315 TTTCTCTCTCTTCTTCTTGGTTCACATATATGCTTAATGACAGTTCATGTCCTACGTCCT 1374
DB |||||
QY 753 GCCTTCGAATGAGGAGCATGTATACCCAGGTACA 786
DB |||||
QY 1375 GCCTTCGAATGAGGAGCATGTATACCCAGGTACA 1408
DB |||||

RESULT 11

ACN39490
ID ACN39490 standard; cDNA; 620 BP.

XX ACN39490;

DT 18-NOV-2004 (first entry)

XX Tumour-associated antigenic target (TAT) cDNA DNA325701, SEQ ID NO:3662.

XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX central nervous system cancer; bladder cancer; pancreatic cancer;
XX cervical cancer; melanoma; leukaemia; hybridisation probe;
XX chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic; gene; ss.

OS Homo sapiens.

XX WO2004030615-A2.

XX 15-APR-2004.

XX 29-SEP-2003; 2003WO-US028547.

XX 02-OCT-2002; 2002US-0414971P.

XX (GETH) GENENTECH INC.

XX Wu TD, Zhang Z, Zhou Y;

XX WPI; 2004-347921/32.

XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.

XX Claim 1; SEQ ID NO 3662; 7273pp; English.

XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acids and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT nucleic acid of the invention

SQ Sequence 620 BP; 189 A; 107 C; 154 G; 170 T; 0 U; 0 Other;

Query Match 69.1%; Score 603; DB 13; Length 620;
Best Local Similarity 100.0%; Pred. No. 4.2e-157;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GGCAGCAGAGAGAGAAAGTGGTGAATAATTACATCTGAAATACCAACAGACTGAGAGATGCA 330
DB |||||
QY 1 GGCAGCAGAGAGAGAAAGTGGTGAATAATTACATCTGAAATACCAACAGACTGAGAGATGCA 60
DB |||||

QY 331 GAAGAGGCTGAACGATTCAATGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGGC 390
Db 61 GAAGAGGCTGAACGATTCAATGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGGC 120
QY 391 AGCTAGGTTGGGATTTCTTCAGTTCGAAACAAAAGGCTGTCATCTGATACAAACCTAT 450
Db 121 AGCTAGGTTGGGATTTCTTCAGTTCGAAACAAAAGGCTGTCATCTGATACAAACCTAT 180
QY 451 GGTAACTTGGATTAAGCTGAAGGAAGAGCTCAAGATTTGGTTTGAATGCTCTCAAT 510
Db 181 GGTAACTTGGATTAAGCTGAAGGAAGAGCTCAAGATTTGGTTTGAATGCTCTCAAT 240
QY 511 CTCAGAAAGTCTGAAGATGATGAGAACTCAAAAGAGGAGGCGGATTTGGGATTTGT 570
Db 241 CTCAGAAAGTCTGAAGATGATGAGAACTCAAAAGAGGAGGCGGATTTGGGATTTGT 300
QY 571 CACAAGTTCACTGGAATCTGAACACACAGAGGATACAGAGGCAAGAGAGGAAAGAGC 630
Db 301 CACAAGTTCACTGGAATCTGAACACACAGAGGATACAGAGGCAAGAGAGGAAAGAGC 360
QY 631 AGAGGCTTTGGGATTTGCTGATGAAGATTCCTGATATTTCTGTTCTCCAGTGTTC 690
Db 361 AGAGGCTTTGGGATTTGCTGATGAAGATTCCTGATATTTCTGTTCTCCAGTGTTC 420
QY 691 CATTTCTCTCTCTCTCTCTCTCTCAATATATGCTCAATGACAGTCAATGCTACGTC 750
Db 421 CATTTCTCTCTCTCTCTCTCTCTCAATATATGCTCAATGACAGTCAATGCTACGTC 480
QY 751 CTGCTCGCAATGAGGAGCATGTACCCAGGTATCATCCATGAACCTGGGAGGATTTG 810
Db 481 CTGCTCGCAATGAGGAGCATGTACCCAGGTATCATCCATGAACCTGGGAGGATTTG 540
QY 811 ACTTATTGCTGTTTCAAGTTTAAAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 870
Db 541 ACTTATTGCTGTTTCAAGTTTAAAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 600
QY 871 AAT 873
Db 601 AAT 603

RESULT 12

ADL62461/C
ID ADL62461 standard; DNA; 2553 BP.

XX AC ADL62461;

XX DT 20-MAY-2004 (first entry)

XX DE Human ovarian cancer DNA marker #20673.

XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.

XX OS Homo sapiens.

XX PN WO200170979-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US0009126.

XX PR 21-MAR-2000; 2000US-0191031P.

XX PR 25-MAY-2000; 2000US-0207124P.

XX PR 15-JUN-2000; 2000US-0211940P.

XX PR 07-JUL-2000; 2000US-0216820P.

XX PR 25-JUL-2000; 2000US-0220661P.

XX PR 21-DEC-2000; 2000US-0257672P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Lee J, Lillie J;

DR WPI; 2001-611502/70.

XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian cancer cells as compared to their normal non-cancerous ovarian cells are used to characterize stage, grade, histological type of ovarian cancer.

XX Disclosure; SEQ ID NO 20673; 106pp; English.

XX The invention relates to nucleic acid markers which are overexpressed in ovarian cancer cells as compared to their expression in normal (i.e. non-cancerous) ovarian cells. The invention also relates to polypeptides encoded by the markers, antibodies that selectively bind to the polypeptides, a method of inhibiting ovarian cancer in a patient at risk of developing ovarian cancer involving inhibiting expression of a gene corresponding to a marker of the invention and a method of treating a patient afflicted with ovarian cancer comprising providing to cells of the patient an antisense oligonucleotide complementary to a marker of the invention. The markers are useful for assessing if a patient is afflicted with ovarian cancer, which involves comparing the level of expression of the marker in a patient sample and a normal level of expression of the marker in a control non-ovarian cancer sample. A difference between the expression levels indicates ovarian cancer. The level of expression of a marker corresponds to a secreted protein or to a transcribed polynucleotide or its portion. The level of expression of the marker is assessed by detecting the presence in the sample, a protein or protein fragment corresponding to the marker. The presence of protein or protein fragment is detected using an antibody that specifically binds with the protein or protein fragment. Alternatively, the level of expression of the marker is assessed by detecting the presence of a transcribed polynucleotide which anneals with the marker or anneals with a portion of the polynucleotide comprising the marker, under stringent conditions. The marker is also used for monitoring the progression of ovarian cancer in a patient which involves detecting expression of the marker in a patient sample at a first point in time, repeating the method at a subsequent time and comparing the level of expression. The method is carried out using an ovarian tissue sample. A composition comprising a marker, polypeptide or antibody of the invention is used to treat ovarian cancer. This sequence represents a human ovarian cancer DNA marker of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 2553 BP; 690 A; 597 C; 570 G; 696 T; 0 U; 0 Other;

XX Query Match 68.2%; Score 595.2; DB 5; Length 2553;

XX Best Local Similarity 96.9%; Pred. No. 1.2e-154;

XX Matches 628; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

QY 202 GGAAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTG-AAAAA 260

Db 695 GGAAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTGAAAAA 636

QY 261 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTTACATCTGAATACACAGACTG 320

Db 635 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTTACATCTGAATACACAGACTG 576

QY 321 AGAGAAATGCAGAGAGGGCTGAACGATTCATCTACCTGTGAGCTTGGAGATGAAGAAG 380

Db 575 AGAGAAATGCAGAGAGGGCTGAACGATTCATCTACCTGTGAGCTTGGAGATGAAGAAG 516

QY 381 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGCTGTCTCATCTGATA 440

Db 515 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGCTGTCTCATCTGATA 456

QY 441 ACAAACCTATGTTTAACTTTGGATAAGCTGAAGGAAAGAGCTCAAGAGATTTGGTTGAATG 500

Db 455 ACAAACCTATGTTTAACTTTGGATAAGCTGAAGGAAAGAGCTCAAGAGATTTGGTTGAATG 396

QY 501 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAGGAGCGAT 560

Db 395 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAGGAGCGAT 336

QY 561 TTGGATTTGTCAAGTTTCAGCTGGAACCTGAAACACAGAGATACAGAGCAAGAGA 620

Db 335 TTGGGATTTGTCACAAAGTTACAGTGGAACTGGAACCCACAGAGGATACAGAGGCAAGAGA 276
Qy 621 GGAAGAGAGAGAGCGCTTTGGGATTCCTGATGAAAGTTCTGATACCTTCTGTTCTC 680
Db 275 GGAAGAGAGAGAGCGCTTTGGGATTCCTGATGAAAGTTCTGATACCTTCTGTTCTC 216
Qy 681 CAGTGTTCCTCATTTCT 740
Db 215 CAGTGTTCCTCATTTCT 156
Qy 741 TGCTAGCTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACTGCGG 800
Db 155 TGCTAGCTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACTGCGG 96
Qy 801 CAGAGTTTGAATGCTTGTCTTTCAGCTTTAAGGTTGTTGTTT 848
Db 95 GCAGCAGTTGACTTA-TGCTGTTTTCAGCTTTAAGGTTGTTGTTT 49

RESULT 13

AD122423
ID AD122423 standard; DNA; 1022 BP.

XX AC AD122423;

XX 22-APR-2004 (first entry)

XX Rat liver differentially expressed cDNA seq id 233.

XX hepatotropic; antiinflammatory; virucide; cytostatic; gene therapy;
XX differentially gene expression; liver; toxin; liver disorder;
XX biliary cirrhosis; X-linked adrenoleukodystrophy; Zellweger syndrome;
XX hepatorenal syndrome; hepatitis; hepatocarcinoma; metabolic response;
XX toxicological response; ss; EST; expressed sequence tag; rat.

XX Rattus sp.

XX US2003165854-A1.

XX 04-SEP-2003.

XX 05-DEC-2001; 2001US-00006285.

XX 05-DEC-2000; 2000US-0251986P.

XX (CUNN/) CUNNINGHAM M J.

XX (KASE/) KASER M R.

XX Cunningham MJ, Kaser MR;

XX WPI; 2003-863697/80.

XX New combination comprising a number of cDNAs that are differentially
PT expressed in a liver treated with a toxin, useful for diagnosing, staging
PT or treating liver disorders (e.g. biliary cirrhosis, hepatitis or
PT hepatocarcinoma).

XX Claim 1; SEQ ID NO 233; 28pp; English.

XX The invention describes a combination comprising a number of cDNAs that
CC are differentially expressed in a liver treated with a toxin and are
CC selected from any of the 514 cDNAs listed in the specification, or their
CC complements. The combination is useful in diagnosing, staging or treating
CC a liver disorder (e.g. biliary cirrhosis, X-linked adrenoleukodystrophy,
CC Zellweger syndrome, hepatorenal syndrome, hepatitis or hepatocarcinoma),
CC in monitoring diagnostic and therapeutic applications, in detecting
CC metabolic and toxicological responses, and in elucidating drug mechanism
CC of action. This sequence represents a cDNA differentially expressed in
CC liver tissues in response to treatment with a toxin.

XX Sequence 1022 BP; 309 A; 160 C; 229 G; 250 T; 0 U; 74 Other;

Query Match 57.1%; Score 498.4; DB 10; Length 1022;
Best Local Similarity 87.3%; Pred. No. 6.4e-128;
Matches 580; Conservative 0; Mismatches 81; Indels 3; Gaps 3;
Qy 52 AAAGCTTGCCGAACATAAGCAAGAATGCTTCTGCTGCTGTTGGAGCAACAAGGAATAAA 111
Db 104 AAAGCTTGCTGAACATAAGCAAGAATGCTTCTGCTGCTGTTAGAGCAACAAGGAATAAA 163
Qy 112 GCAAGATCTTATCCACAGACT-CCAGGCCATATCTTGAAGACATGCTGAAGGAGGCA 170
Db 164 ACAAGATCTTATCAATAGGCTTACAGGCCATATCTTGAAGACATGCTGAAGGAGGCA 223
Qy 171 ATGAAGAAAGATCTACTTGGGAGATGAAACAGAGGAGAAAGAAACAAAGCCATTGAGCTCC 230
Db 224 ATGAAGAAAGATCTACTTGGGAGATGAAACAGAGGAGAAAGAAACAAAGCCATTGAGCTCC 283
Qy 231 CTGTCAAAGAGAAAGAACCCCTCTGAAAAAATCTGTTGATGTGGCAGCAGAGAAAGAGTGG 290
Db 284 CTGTTAAAGAGGAAAGAACCCCTCTGAAAAAAGTTGTTGATATGGCATCAGAAAAAGAGTGG 343
Qy 291 TGAATAATTACATCTGAAATACACAGACTGAGAGATGCGAGAGGGCTGAGACGATTCA 350
Db 344 TAAAAATTACATCTGGAATACCTCAAACTGAGAGATGCGAGAGGGCTGAGACGTTTCA 403
Qy 351 ATGTACCTGTGAGCTTGGAGAGTAAAGAAAGCTGCTCG-GECCAGCTAGGTTTGGGATTTCT 409
Db 404 ATGTGCTGTAAAGCTTGGAGAGTAAAGAGCTGCTCGCGGCGAGAGGTTTGGAAATTTCT 463
Qy 410 TCAGTTCCAAACAAAAGGTTCTGTCATCTGTATACAAACCTATGTTAACTTGAATAAGCTG 469
Db 464 TCAGTTCCAAACAAAAGGTTTATCATCTGACACCAAGCCAAATGGTTAACTTGAATAACTA 523
Qy 470 AAGGAAAGAGCTCAAGAGTTTGGTTGATGTTCTTCAATCTCCAGAAAGTCTGAAGAT 529
Db 524 AAGGAAAGAGCAGAGATTTGGTTTGAATGTTCTTCCATCTCTAAGAAAGTCTGAGGAT 583
Qy 530 GATGAGAAACTGAAAAAGAGAGGAGCGATTGTTGGATTGTCAAGTTTCAAGCTGGAAC 589
Db 584 GATGAGAGCTGAGAAACGGAAGAGAGATTGTTGATTTGTGCAAGTTTCAAGCTGGAAC 643
Qy 590 GGAACCAACAGAGGATACAGAGGCAAGAGAAAGAGGAAAGAGAGCGCTTTGGGATTGCC 649
Db 644 GGAACCAACAGAGGATACAGAGGCAAGAGAAAGAGAAAGAGAGCGCTTTGGGATTGCA 703
Qy 650 TGATGAAAGTTCCCTG-ATACCTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCT 708
Db 704 TAATGAAAGTTCTGCTTTCTGCCCCACCCCATAGTGGTTTCCATTTCTCAGATTTCTT 763
Qy 709 TGGT 712
Db 764 GGT 767
RESULT 14
AAC00738
ID AAC00738 standard; cDNA; 471 BP.
XX AAC00738;
AC AAC00738;
XX 06-OCT-2000 (first entry)
DT Human secreted protein 5' EST, SEQ ID NO: 736.
XX Human secreted protein 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
OS Homo sapiens.
XX EP1033401-A2.
PN EP1033401-A2.
XX 06-SEP-2000.
PD 21-FEB-2000; 2000EP-00200610.
XX PF

```
XX 26-FEB-1999; 99US-0122487P.
XX (GEST ) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX P-PSDB; AAG00732.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 1; SEQ ID NO 736; 71pp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. An ORF has been identified within the
XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
XX derived from 30 different tissues. EST sequences usually correspond
XX mainly to the 3' untranslated region (UTR) of the mRNA because they are
XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
XX well suited for isolating cDNA sequences derived from the 5' ends of
XX mRNAs and even in those cases where longer cDNA sequences have been
XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
XX mRNAs with intact 5' ends and can therefore be used to obtain full length
XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
XX gene therapy and chromosome mapping procedures. They are used to obtain
XX upstream regulatory sequences and to design expression and secretion
XX vectors
XX
XX SQ Sequence 471 BP; 165 A; 81 C; 130 G; 94 T; 0 U; 1 Other;
Query Match 53.3%; Score 465.6; DB 3; Length 471;
Best Local Similarity 99.8%; Pred. No. 6e-119;
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGGAGTGGAGGTTAAAGATGCGACCGAGCGGTGGAGTCCATAGCTTAAAGCTTGC 60
Db 6 TGGAGTGGAGGTTAAAGATGCGACCGAGCGGTGGAGTCCATAGCTTAAAGCTTGC 65
Qy 61 CGAACTAAAGCAAGATGCTTGTCTGTGTTGGAGACCAAGGGAATAAGCAAGATCT 120
Db 66 CGAACTAAAGCAAGATGCTTGTCTGTGTTGGAGACCAAGGGAATAAGCAAGATCT 125
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGACATGCTGAAGAGGAGGCAATGAAGA 180
Db 126 TATCCACAGACTCCAGGCATATCTTGAAGACATGCTGAAGAGGAGGCAATGAAGA 185
Qy 181 TGTAATGGAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Db 186 TGTAATGGAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 245
Qy 241 GGAAGAACCCCTGAAAGAACTGTTGATGTGGAGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 246 GGAAGAACCCCTGAAAGAACTGTTGATGTGGAGAGGAGGAGGAGGAGGAGGAGGAGG 305
Qy 301 ATCTGAATACCAAGAGAGTGAAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
Db 306 ATCTGAATACCAAGAGAGTGAAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 365
Qy 361 GAGCTTGGAGAGTGAAGAGGAGTGGCGGAGGAGTGGGAGTGGGAGTGGGAGTGGGAGG 420
Db 366 GAGCTTGGAGAGTGAAGAGGAGTGGCGGAGGAGTGGGAGTGGGAGTGGGAGTGGGAGG 425
Qy 421 AAAAGGCTGTGCTATCTGATGAACAAACCTATGGTTAACTTGGATAAG 466
Db 426 AAAAGGCTGTGCTATCTGATGAACAAACCTATGGTTAACTTGGATAAG 471
RESULT 15
ADP28791
ID ADP28791 standard; DNA; 558 BP.
```

```
XX ADP28791;
XX AC
XX 12-AUG-2004 (first entry)
XX DE Human secreted protein encoding sequence SEQ ID #789.
XX DE Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
XX KW cancer; inflammatory; immune; ds; human secreted protein.
XX OS Homo sapiens.
XX PN WO2004035732-A2.
XX PD 29-APR-2004.
XX
XX 28-AUG-2003; 2003WO-US026780.
XX 29-AUG-2002; 2002US-0406576P.
XX 29-AUG-2002; 2002US-0406579P.
XX 29-AUG-2002; 2002US-0406585P.
XX 29-AUG-2002; 2002US-0406588P.
XX 29-AUG-2002; 2002US-0406608P.
XX 29-AUG-2002; 2002US-0406611P.
XX 29-AUG-2002; 2002US-0406612P.
XX 29-AUG-2002; 2002US-0406616P.
XX 29-AUG-2002; 2002US-0406640P.
XX 29-AUG-2002; 2002US-0406642P.
XX 29-AUG-2002; 2002US-0406646P.
XX 29-AUG-2002; 2002US-0406653P.
XX 29-AUG-2002; 2002US-0406655P.
XX 29-AUG-2002; 2002US-0406666P.
XX 17-SEP-2002; 2002US-0410946P.
XX 17-SEP-2002; 2002US-0410947P.
XX 17-SEP-2002; 2002US-0410948P.
XX 17-SEP-2002; 2002US-0410949P.
XX 17-SEP-2002; 2002US-0410953P.
XX 17-SEP-2002; 2002US-0410957P.
XX 17-SEP-2002; 2002US-0410958P.
XX 17-SEP-2002; 2002US-0410959P.
XX 17-SEP-2002; 2002US-0410960P.
XX 17-SEP-2002; 2002US-0410961P.
XX 17-SEP-2002; 2002US-0410962P.
XX 17-SEP-2002; 2002US-0411019P.
XX 17-SEP-2002; 2002US-0411022P.
XX 17-SEP-2002; 2002US-0411023P.
XX 17-SEP-2002; 2002US-0411024P.
XX 17-SEP-2002; 2002US-0411032P.
XX 17-SEP-2002; 2002US-0411035P.
XX 17-SEP-2002; 2002US-0411037P.
XX 17-SEP-2002; 2002US-0411041P.
XX 17-SEP-2002; 2002US-0411045P.
XX 17-SEP-2002; 2002US-0411046P.
XX 17-SEP-2002; 2002US-0411048P.
XX 17-SEP-2002; 2002US-0411052P.
XX 17-SEP-2002; 2002US-0411055P.
XX 17-SEP-2002; 2002US-0411073P.
XX 17-SEP-2002; 2002US-0411082P.
XX 17-SEP-2002; 2002US-0411101P.
XX 17-SEP-2002; 2002US-0411111P.
XX 18-APR-2003; 2003US-0463700P.
XX 18-APR-2003; 2003US-0463708P.
XX 18-APR-2003; 2003US-0463716P.
XX 18-APR-2003; 2003US-0463732P.
XX 02-MAY-2003; 2003US-0467199P.
XX 02-MAY-2003; 2003US-0467201P.
XX 02-MAY-2003; 2003US-0467203P.
XX 02-MAY-2003; 2003US-0467230P.
XX 19-MAY-2003; 2003US-0471306P.
XX 19-MAY-2003; 2003US-0471336P.
XX 22-MAY-2003; 2003US-0472420P.
XX 22-MAY-2003; 2003US-0472430P.
XX 09-JUN-2003; 2003US-0476609P.
```


